The Dissertation Committee for Yun Huang Certifies that this is the approved version of the following dissertation:

ECONOMIC ISSUES IN DISTRIBUTED COMPUTING

Committee:

Andrew B. Whinston, Supervisor

Kenneth Hendricks

Gary J. Koehler

John Mote

Gautam Ray

ECONOMIC ISSUES IN DISTRIBUTED COMPUTING

by

Yun Huang, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2007

Dedication

To Yuning and Kayla, for always supporting my endeavors.

Acknowledgements

First and foremost, I would like to thank my advisor and dissertation chairman, Andrew B. Whinston who played an indispensable role in my intellectual development. I am truly grateful for his thoughtful insights into research as well as the invaluable constant support along the difficult trail of finishing my dissertation. I am also grateful for the other members of my dissertation committee: Dr. Kenneth Hendricks, Dr. Gary J. Koehler, Dr. John Mote, and Dr. Gautam Ray, for their comments on portions of this dissertation.

I thank my coauthors: Dr. Xianjun Geng, Dr. Bin Gu, Dr. Wenjing Duan, and the fellow students from the CREC research center. They have provided me great guidance, support, and advice on my research in general.

Finally, I am sincerely grateful to my wife Yuning and daughter Kayla for their love and support which helped me overcome all hurdles in finishing my PhD study.

Economic Issues in Distributed Computing

Publication No._____

Yun Huang, Ph.D.

The University of Texas at Austin, 2007

Supervisor: Andrew B. Whinston

On the Internet, one of the essential characteristics of electronic commerce is the integration of large-scale computer networks and business practices. Commercial servers are connected through open and complex communication technologies, and online consumers access the services with virtually unpredictable behavior. Both of them as well as the e-Commerce infrastructure are vulnerable to cyber attacks. Among the various network security problems, the Distributed Denial-of-Service (DDoS) attack is a unique example to illustrate the risk of commercial network applications. Using a massive junk traffic, literally anyone on the Internet can launch a DDoS attack to flood and shutdown an eCommerce website.

Cooperative technological solutions for Distributed Denial-of-Service (DDoS) attacks are already available, yet organizations in the best position to implement them lack incentive to do so, and the victims of DDoS attacks cannot find effective methods to motivate the organizations. Chapter 1 discusses two components of the technological solutions to DDoS attacks: cooperative filtering and cooperative traffic smoothing by caching, and then analyzes the broken incentive chain in each of these technological

solutions. As a remedy, I propose usage-based pricing and Capacity Provision Networks, which enable victims to disseminate enough incentive along attack paths to stimulate cooperation against DDoS attacks.

Chapter 2 addresses possible Distributed Denial-of-Service (DDoS) attacks toward the wireless Internet including the Wireless Extended Internet, the Wireless Portal Network, and the Wireless Ad Hoc network. I propose a conceptual model for defending against DDoS attacks on the wireless Internet, which incorporates both cooperative technological solutions and economic incentive mechanisms built on usage-based fees. Cost-effectiveness is also addressed through an illustrative implementation scheme using Policy Based Networking (PBN). By investigating both technological and economic difficulties in defense of DDoS attacks which have plagued the wired Internet, our aim here is to foster further development of wireless Internet infrastructure as a more secure and efficient platform for mobile commerce.

To avoid centralized resources and performance bottlenecks, online peer-to-peer communities and online social network have become increasingly popular. In particular, the recent boost of online peer-to-peer communities has led to exponential growth in sharing of user-contributed content which has brought profound changes to business and economic practices. Understanding the dynamics and sustainability of such peer-to-peer communities has important implications for business managers. In Chapter 3, I explore the structure of online sharing communities from a dynamic process perspective. I build an evolutionary game model to capture the dynamics of online peer-to-peer communities. Using online music sharing data collected from one of the IRC Channels for over five years, I empirically investigate the model which underlies the dynamics of the music sharing community. Our empirical results show strong support for the evolutionary process of the community. I find that the two major parties in the community, namely

sharers and downloaders, are influencing each other in their dynamics of evolvement in the community. These dynamics reveal the mechanism through which peer-to-peer communities sustain and thrive in a constant changing environment.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: DEFEATING DDoS ATTACKS BY FIXING THE INCENTIVE CHAIN

Cooperative technological solutions for Distributed Denial-of-Service (DDoS) attacks are already available, yet organizations in the best position to implement them lack incentive to do so, and the victims of DDoS attacks cannot find effective methods to motivate them. In this paper we discuss two components of the technological solutions to DDoS attacks: cooperative filtering and cooperative traffic smoothing by caching. We then analyze the broken incentive chain in each of these technological solutions. As a remedy, we propose usage-based pricing and Capacity Provision Networks, which enable victims to disseminate enough incentive along attack paths to stimulate cooperation against DDoS attacks.

"More than a dozen offshore gambling sites serving the US market were hit by the so-called Distributed Denial of Service (DDoS) attacks and extortion demands in September. Sites have been asked to pay up to $50,000 (€43,000, £30,000) to ensure they are free from attacks for a year. Police are urging victims not to give in to blackmail and to report the crime." – Chris Nuttall, The Financial Times, November 12, 2003, Front page – First section.

INTRODUCTION

Internet-enabled business, or e-business, has mushroomed into a significant part of the US economy, yet further advancement of e-business is plagued by various Quality-of-Service (QoS) and security problems. One of the worst is the Distributed Denial-of-Service (DDoS) attack, which aggregates junk data traffic from up to thousands of computers into a formidable volume and floods and effectively blocks a certain victim website. DDoS attacks have drawn a lot of media attention since the landmark attacks on

a large portfolio of famous e-business websites including Yahoo!, Amazon, CNN, eBay, and E*Trade in early 2000 (Kleinbard 2000). Cavusoglu et al. (2002) estimate that the firms involved lost more than 2.8% of their market capitalization. Academic discussion also quickly followed up with proposals that can be broadly classified into two categories: technological solutions (Wang and Reiter 2004, Badishi, Keidar, and Sasson 2004, Xiang, Zhou, and Chowdhury 2004, Mirkovic et al. 2005 Chapter 7), and economic solutions (Geng and Whinston 2000, Geng, Huang and Whinston 2002).

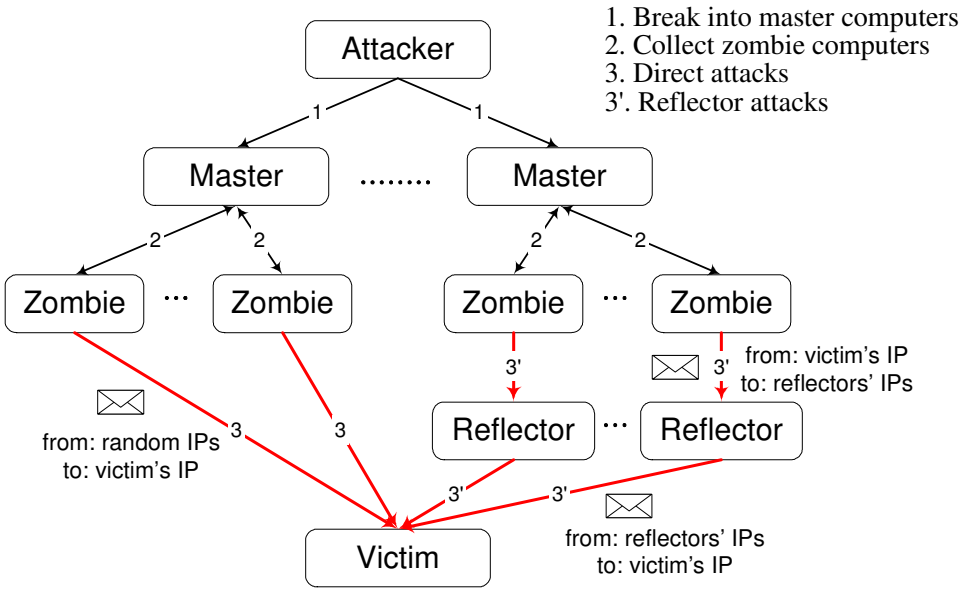

Figure 1.1. The mechanism of DDoS attacks

Figure 1.1 illustrates the mechanisms of a DDoS attack. There are two separate stages of DDoS attacks: recruiting zombies and flooding the victim (Chang 2002). In the recruiting stage (steps 1 and 2), security flaws are used to break into master computers and a large set of zombie computers is established. In the flooding stage, a direct attack or a reflector attack is launched and synchronized traffic with IP spoofing (Geng and Whinston 2000) disables the services of the victim (steps 3 and 3').

It is now well-understood that several cooperative technological solutions including cooperative filtering and cooperative traffic smoothing by caching (as we will shortly discuss) will be quite effective against DDoS attacks *if* they are implemented. Nevertheless, it turns out to be a big "if": while some of those cooperative technological solutions were proposed as early as in 2000 (e.g. RFC 2827 - ubiquitous ingress filtering), they are clearly not effectively deployed since DDoS attacks still threaten the Internet (Naraine 2002), sometimes in a more hazardous way: in one recent incidence, a Russian mafia launched DDoS attacks and brought down more than a dozen offshore gambling sites serving the US market, such as www.VIPSports.com and www.Betgameday.com (The Financial Times, November 12, 2003). The gangs blackmailed the victims and demanded $50,000 for one-year protection.

On the surface, this seems to be an irony: The consensus is that defeating DDoS attacks will be beneficial to e-business giving the huge lost these attacks cost; however, organizations are still reluctant to establish the defense. For example, despite some occasional practices, cooperative filtering is still not widely used on the Internet.

In this paper we argue that, although there is room for more improvements in technological solutions, the priority should be placed on economic solutions given the extremely unbalanced development in these two directions: until now a vast amount of research has been done on technological solutions while only a handful exist on economic parts. In fact, the irony is just the result of the continuing ignorance of incentive issues involved in dealing with DDoS attacks, and the intuition behind this irony is actually quite straightforward: *the parties that suffer the most are not in the best position to defend, while the parties in the best position do not suffer enough to defend.* Successful delivery of digital contents requires the collaboration of multiple parties, such as Internet Content Providers (ICP), Backbone ISPs, regional ISPs, and end users. Each

of them makes independent decisions to invest and contribute to the delivery process. The value of final products depends on the effort of each party. An *incentive chain*, which is the set of value and monetary transactions along digital delivery channels, acts as a glue to stick various parties together in collaboration. In case of DDoS attacks the incentive chain is often broken: defensive actions by ISPs benefit ICPs and end users the most, yets ISPs are rarely compensated for – and thus often under-motivated to take – these actions.

How to fix the broken incentive chain? Or phrased differently: how to transfer incentives from the parties that suffer the most to the parties that are in the best position to defend? We propose an integrated framework that has two pillars: a usage-based traffic pricing structure that stimulates cooperative filtering (Geng and Whinston 2000), and a market for demand-side cache trading called Capacity Provision Networks (CPN) that gives cache owners incentives to perform cooperated traffic smoothing (Geng et al. 2003, 2005).

The rest of the paper is organized as follows. Section 2 discusses the structure of digital supply chains and explains what effective technological solutions are readily available against DDoS attacks. Section 3 analyzes the incentive chain on the Internet, and shows why it is broken under DDoS attacks, and thus why the technological solutions fail to materialize. Section 4 illustrates the remedies and associated costs. In Section 5, we discuss some key players and technologies – especially overlay networks and Internet mapping – that have significant impacts on potential remedies. Section 6 concludes this paper.

THE DIGITAL SUPPLY CHAIN AND COOPERATIVE TECHNOLOGICAL SOLUTIONS TO DDOS ATTACKS

The Digital Supply Chain

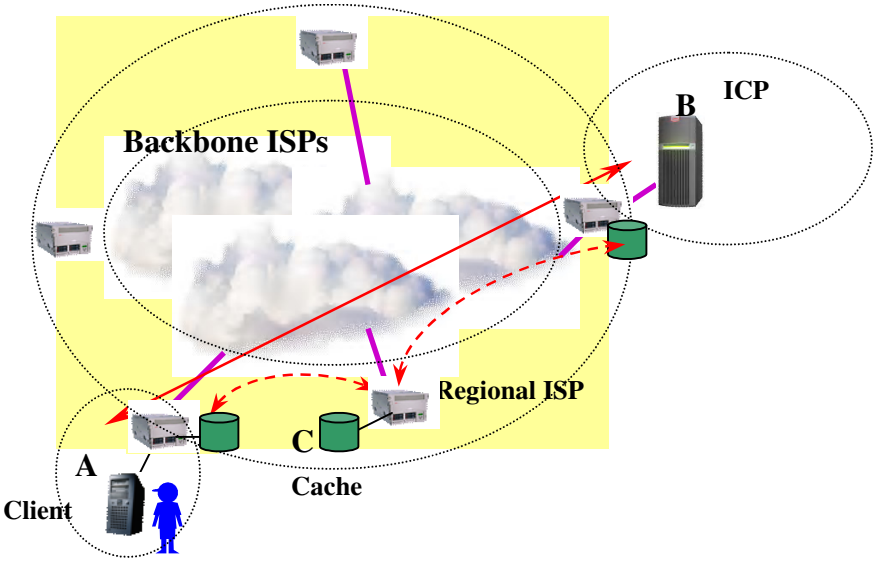

Figure 1.2. The digital supply chain

To understand how the technological solutions against DDoS attacks work, we need to first understand how the Internet-based digital supply chain works. Figure 1.2 gives a simple illustration of the Internet structure, which consists of the following components (data from navigators.com):

1. The Internet core, which consists of dozens of interconnected backbone ISPs who collectively maintain the backbone of the Internet.
2. The Internet cloud except the core, which consists of less than 10,000 regional ISPs that connect to the core through one or several backbone ISPs and serve different geographical regions.
3. The edge of the Internet, which consists of around 100,000 networks that are locally administrated.
4. Millions of online computers including content servers and clients.

Besides transmitting data packets, regional ISPs and local networks often provide caching, which is the temporary storage of data for quick retrieval by local users.

In this structure, content requests and delivery are fulfilled in a variety of ways including:

1. Direct communication between clients and content servers, as the route from client A to server B in Figure 1.2.
2. Cache-intermediated communication where a cache server first stores a copy of server's content, which is then accessed by adjacent clients, as the route from client A to cache C in Figure 1.2.

With respect to a commercial Internet content provider (ICP), not all the content requests are equally valuable: if the ICP is an online retailer, the requests that lead to purchases are most valuable. Nevertheless, the current Internet infrastructure does not distinguish between requests. When congestion happens, some requests have to be dropped, and the ones dropped could be valuable. In the scenario of DDoS attacks, as a content server is jammed by junk requests, almost all valuable requests from legitimate clients are dropped, which leads to a significant financial loss and reputation damage for the ICP.

Cooperative Filtering

When it comes to defending against DDoS attacks, there are several common misunderstandings. One common misunderstanding is that "we still do not have any effective technological solution against DDoS attacks." We do. The first one is cooperative filtering, where the ISPs along attack paths cooperatively filter out attack traffic (Geng and Whinston 2000). The second one is cooperative caching, where instead of filtering out attack traffic, ISPs try to divert it to multiple caches such that each cache (and the ICP) only bears a manageable fraction of the whole traffic (Naraine 2002). The details of both are explained shortly. Another common misunderstanding is that "the victim of a DDoS attack can defend itself by implementing some security and/or traffic control systems in its own boundary." The victim cannot do it efficiently. Although intrusion detection systems on the victim side can prevent attack traffic from reaching

content servers, DDoS attacks can simply flood the entrance (i.e. the router) to its network, rendering all local defenses irrelevant. One direct implication is that technological solutions can be effective only if other involved parties cooperate.

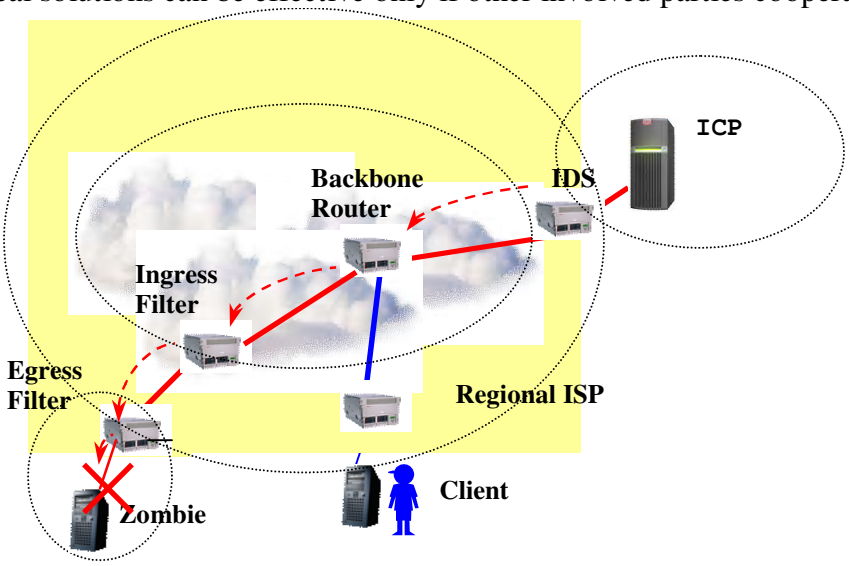

Figure 1.3. The process of cooperative filtering

Cooperative filtering is the first cooperative technological solution. Cooperative filtering works in three steps: alarming, tracing, and filtering (illustrated in Figure 1.3). By analyzing the pattern of network traffic, Intrusion Detection Systems (IDS) identify suspicious traffic and send out alarms. Following the alarms, a tracing mechanism kicks in to track back each attack path as far as possible. Finally, a series of filters along every attack path are configured to filter out attack traffic. In the best scenario, a tracing mechanism may find the computers (zombies) that are initiating attack traffic, and may inform the responsible ISPs to take them offline.

One simple approach of cooperative filtering is to ban IP-spoofing at the edge of the Internet. Attack traffic consists of a large amount of IP packets, where each packet contains its source-address. If source addresses are correct, the tracing mechanism will be very simple: just informing the ISP serving that IP address to take that computer offline.

Nevertheless, IP-spoofing is the technology which modifies the source-address and makes it useless. Note that IP-spoofing can be detected by the immediate ISPs of attacking computers: an ISP can simply compare the source-address with the route the packet comes from – if they are not consistent, IP-spoofing is detected. Therefore, if IP-spoofing is banned everywhere at the edge of the Internet, tracing DDoS attack sources will be quite straightforward.

Cooperative Caching

Instead of filtering out attack traffic, DDoS attacks can also be defeated by diverting and evenly distributing attack traffic from a victim into a large number of cache servers such that each stream of diverted traffic is not significant enough in volume to create any congestion (Geng et al. 2003), as shown in Figure 1.5.

Cooperative caching is an effective solution to DDoS attacks when cooperative filtering is costly to implement, or when attack traffic is well concealed in legitimate data requests such that pattern recognition is technically difficult. Cooperative caching and filtering can also be jointly deployed so that attack traffic is both reduced and diverted, resulting in a more effective defense.

One important technological issue using cooperative caching to defeat DDoS attacks is the fact that only relatively static content can be cached. If a DDoS attack targets dynamic content or protocols (such as ICMP ECHO, SYN floods, BGP floods), traditional caching solutions cannot divert it. This issue is now partially addressed in two ways. First, standards like Edge Side Include (ESI, see www.esi.org) enable caching of dynamic content. Second, more ISPs start screening and restricting control packets. For example, the attacks using ping commands are no longer effective when ICMP traffic is restricted.

THE BROKEN INCENTIVE CHAIN

Despite the fact that cooperative filtering and cooperative caching are two effective technological solutions against DDoS attacks, to date they have been rarely deployed on the Internet. The reason becomes quite straightforward once we look at the digital supply chain from another perspective: that of the incentive chain.

There are two major sources driving the flow of digital content on the commercial Internet: end users' demand to consume digital content and ICPs' demand to publish digital content. As shown in Figure 1.4, while both end users and ICPs only pay directly to their ISPs for Internet connections, those regional ISPs in turn pay larger regional ISPs and backbone ISPs for the connectivity to the core of the Internet. We call this series of payments the "incentive chain," which acts as glue to link all parties together in the end-to-end transmission of digital content.

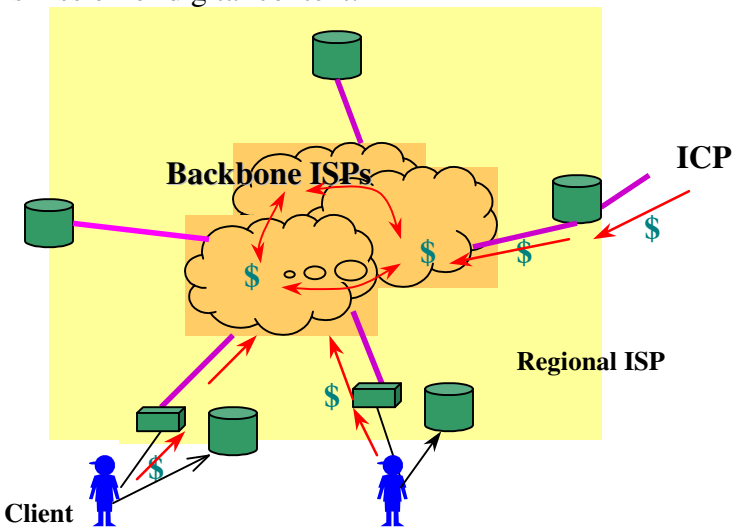

Figure 1.4. The incentive chain

Lack of Incremental Payment Structure and the Failure of Cooperative Filtering

Today the dominant payment method on the Internet is subscription, where an end user/regional ISP pays a fixed monthly fee to a regional ISP/backbone ISP for Internet connections. Within a certain range of traffic volumes that can be accommodated by specific Internet connections, actual usage is irrelevant to monthly payments. As a result, the traffic pattern of individual users is often highly volatile. To avoid congestion and guarantee a certain level of quality of service, ISPs tend to be conservative in planning and acquire extra bandwidth.

One important implication of this conservative practice in uplink planning is that most of the time ISPs have abundant residue bandwidth unused that they already paid for to the upper-level ISPs. ISPs are willing to provide such an unused resource for better consumer retention, and on the surface it appears to not hurt anybody else. However, it actually leads to devastating consequences on cooperative filtering against DDoS attacks once we look at the question: what are the costs and benefits for an ISP to engage in cooperative filtering? While the cost side includes the administrative work in setting up and maintaining filters, and the reduction of transmission performance due to filtering overhead; the benefit side often includes little to nothing as long as DDoS attacks only consume some of the residue bandwidth, which is unused anyway.

As an example, recall the infamous DDoS attacks in early 2000 that disabled websites like Yahoo!. Giving the distributed nature of DDoS attacks, attack traffic from each zombie computer was not huge enough to cause any local congestion and mostly even not noticed by its local ISP. In other words, defending against those DDoS attacks brings little benefit, or none at all, to those local ISPs. Then what are the incentives for those local ISPs to cooperate with Yahoo! by setting up costly filters?

Although subscription is the usual payment method for Internet access, in practice the victims of DDoS attacks may have more flexibility in interacting with their direct ISPs. In the Yahoo! case, given the importance of Yahoo! as a large corporate client, its ISP actually closely cooperated with it to filter out attack traffic. However, such flexibility rapidly diminishes alone the incentive chain – it took Yahoo! days just to persuade ISPs along attack paths to help trace back to zombie computers.

The inability of victims in DDoS attacks to motivate ISPs who are in the best position to filter attack traffic is the direct result of the lack of incremental payment structures on the Internet. By selling and buying Internet access on a subscription basis, ISPs have little incentive to control traffic volumes as long as it does not create congestion in their own neighborhoods, simply because the marginal cost for transmitting additional data packets is zero. Additional bandwidth may be used to initiate DDoS attacks and harm ICPs far away. However, this does not provide any incentive for local ISPs to take any action. Clearly, when it comes to a DDoS attack, the incentive chain is broken.

Caches on the Edge of the Internet: Inaccessible Treasures

The optimization of an incentive chain is all about the tradeoffs between the costs and benefits of various possible incentive schemes. As we noted before, cooperative filtering is actually costly to ISPs because of administrative costs and performance reduction. Alternatively, if DDoS attack traffic can be diverted to a lot of cache servers through cooperative caching, it can be an effective solution as it prevents the accumulation of traffic from happening. Thinking of DDoS attacks as floods in the Mississippi River, then the cache servers distributed over the Internet are like the countless lakes along the Mississippi that absorb the floods. Since cache servers already

exist, as long as cooperative caching only uses redundant cache capacity, it incurs little cost to any party involved, and thus is more cost efficient than cooperative filtering.

Nevertheless, as shown in Figure 1.4, ISPs' caches only serve their local users who pay for connections. Congestion at the ICP's website does not provide any payment for cache servers on the demand side to engage in cooperative caching. Therefore, the resource is inactive in the theme of defending DDoS attacks, and again the incentive chain is broken.

Note that the reasons for broken incentive chains are different in cooperative filtering and in cooperative caching: in the former, the payment structure exists – it just does not offer proper incentives for cooperation; in the latter, there is simply no payment structure. The reason that there is no payment structure from ICPs to regional ISPs who own demand-side caches is straightforward: an ICP is never sure where DDoS attacks will come from, and which cache servers are in the best positions to help. Signing payment contracts with each of tens of thousands of regional ISPs to have a full-fledge defense against possible DDoS attacks is impractical.

FIXING THE INCENTIVE CHAIN

Fixing the incentive chain requires two types of effort. First, in the scenarios where payment structures already exist but do not offer the right incentive, the incentive chain needs to be reengineered accordingly. For instance, a usage-based, instead of a subscription-based, pricing structure provides the right incentive for cooperative filtering. Second, if payment structures do not exist at all, new market mechanisms need to be established. In the context of cooperative caching, we apply the Capacity Provision Network (CPN) market mechanism to enable regional ISPs to collectively offer caching services to ICPs. As Internet mapping is indispensable for cooperative filtering and caching, we also discuss the proper incentive structures for Internet mapping.

Usage-Based Pricing

Usage-based pricing ties payments to actual traffic volumes, thus is an incremental payment structure. One common form of usage-based pricing is IP-packet based pricing, where the cost for transmitting one IP packet is fixed. It is used, for example, by NTT DoCoMo in pricing i-mode services in Japan. Another form of usage-based pricing that is heavily studied in academia is congestion-based pricing, also called dynamic pricing, where the cost for transmitting packets is dynamically adjusted based on network congestion (Gupta, Stahl, and Whinston 1999).

The central requirement of usage-based pricing is that the cost of transmitting the attack traffic generated by zombie computers has to be large enough for related ISPs even when it does not lead to any congestion, thus they have enough incentive to set up filters. As long as this requirement is fulfilled, the specific structure of usage-based pricing is not important in defending against DDoS attacks.

To clarify this statement, we need to point out two aspects of usage-based pricing for defending against DDoS attacks. First, usage-based pricing need not be complex for end users or ISPs to figure out. Take the pricing practice of major US wireless providers as an example: they usually offer a ladder of choices where a customer pays a fixed amount for the minutes within a certain range, and pay per minute fee for any minutes beyond that range. Such a pricing structure offers customers the peace of mind when they are within the range, while also offers them the flexibility to talk more if they want to. This is clearly an example of non-linear usage-based pricing that is quite simple for consumers to calculate their cost. If such a laddered pricing structure is adopted in Internet pricing (just as an example), as long as a DDoS attack leads ISPs to exceed their traffic limits or forces them to pick a higher ladder, it provides ISPs an incentive to filter.

Second, usage-based pricing does not necessarily require the overhead of traffic metering. Due to the nature of the Internet, packet counting is costly and infeasible. Furthermore, since the packets have no explicit value, it is meaningless to charge based on traffic. The "usage" indicates the acquisition of verifiable resources and services, such as bandwidth, cache, and excess traffic (it is a common practice for ISPs). For example, if ISPs can buy extra bandwidth from upper-level ISPs in real time, and DDoS attacks lead them to buy more to mitigate congestion, then these ISPs will have an incentive to filter. Note that in this example, traffic metering is not necessary – it only requires ISPs to be able to dynamically acquire capacity from upper-level ISPs according to network conditions.

Capacity Provision Network

As we mentioned earlier, it is difficult for ICPs to get help from regional ISPs on cooperative caching because signing bilateral contracts with each of them to provide incentives is too costly to be practical. What if all the regional ISPs have their cache capacity organized by an intermediary who in turn deals with ICPs? In this scenario, ICPs only need to deal with and pay a single entity – the intermediary, and the intermediary specializes in dividing the payment and compensating the participating ISPs who provide cache capacity.

Such an intermediary is yet to emerge. In several recent research papers, Geng et al. proposed the concept of a Capacity Provision Network (CPN), which is a network of cache servers owned, operated and coordinated through capacity trading by different ISPs (Geng et al. 2003, 2005). A CPN is initially proposed for demand-side cache trading, the usefulness of which is supported by the fact that there exist positive network externalities across individual ISPs who provide caching services to their respective local users: when some ISPs are experiencing high demand for caching, other ISPs' cache capacity may be

idling. Therefore, by sharing the idling cache capacity with busy ISPs total welfare increases.

Cache trading is operated in a CPN market, which is organized by a market owner. We propose that the owner of the CPN market fits well in the intermediary's role as we described it above: the owner specializes in dealing with large numbers of ISPs who own cache servers, and the owner is a single entity that can deal with outside organizations on behalf of its participating ISPs.

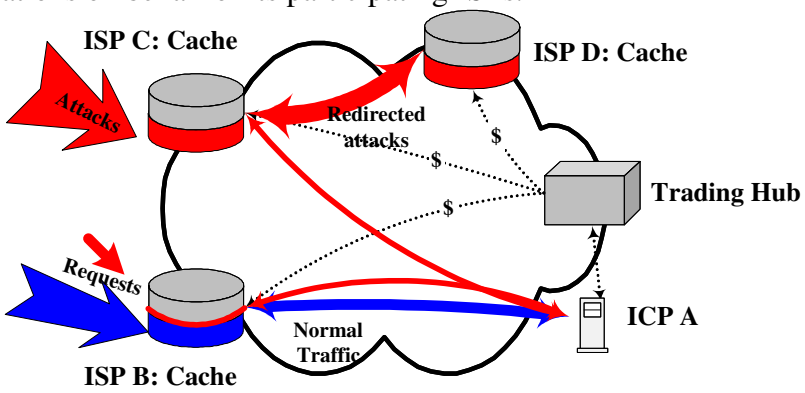

Figure 1.5. Capacity Provision Networks

Figure 1.5 illustrates an incentive chain for a CPN owner intermediated cooperative caching. An ICP initiates the incentive chain by contracting with and paying the CPN owner for cooperative caching against any possible DDoS attacks. When a DDoS attack happens, the CPN owner decides which cache server is in the best position to dilute the traffic and then accordingly pays relevant ISPs to start cooperative caching, which completes the incentive chain. Of course, how much the ICP pays the CPN owner depends in turn on how much the owner pays ISPs.

Note that besides the CPN solution where an intermediary coordinates an army of independent ISPs, a Content Distribution Network (CDN) company can also execute cooperative caching. For example, Akamai provides traffic smoothing for ICPs through

its global network. The operator of a CDN replicates digital content at the edge of networks and redirects requests based on complicated algorithms (Wang, Pai and Peterson 2002). There are three advantages a CPN has over a CDN. First, the collective cache capacity of all regional ISPs is much larger than the capacity of a single CDN company, and is far more distributed. Therefore, a CPN can deliver much better performance with cooperative caching. Second, a CPN needs little initial investment, as it is based on existing caching capacities of regional ISPs, while a CDN requires a considerable initial investment on hardware – there is already abundant cache capacity on the Internet, why build new ones? Finally, the cache capacity traded in a CPN is dedicated to buyers. An ICP can serve local consumers at given time by acquiring capacity and caching specific content. However, in a CDN, since the operator manages the allocation of content and the methods of request redirection, the QoS for each ICP is uncertain. It leads to a major drawback in pricing CDN services: ICPs have to pay in advance for unpredictable evens and service level agreements (SLAs) are difficult to design and verify.

A Simple Model of Cache Trading

To demonstrate the use of CPNs and the differences between CPNs and CDNs, we show an application of CPNs in the scenario of DDoS attacks. Without sophisticated intrusion detection and traffic monitoring, a simple fixed fee for cache capacity can divert traffic away from the server side. As shown in Figure 1.5, when suffering a massive volume of traffic, ICP A can always discover frequently requested content and the major sources of traffic, which could be either legitimate requests or an attack flood. Instead of separating them, ICP A can establish a mirror of popular targets in ISP networks, which directly control the sources of traffic.

A trading hub of the CPN provides a mechanism to use ISP side cache capacity in real-time (Geng et al 2003). A cache server (in cases of ISPs B & C) or remote cache capacity provided by peered ISPs (in ISP D's case) improves the QoS of legal users and blocks a large amount of malicious traffic. Since this solution does not change ISP side redirectors, such as cache redirectors, local filters, and DNS servers, ISPs have no extra investment. A simple static payment for capacity will create incentives for ISPs to share local resources.

In this scenario, an ICP pays for what it uses: buying and utilizing cache servers in a certain time period for specified users. As long as gains in terms of increased QoS or decreased damage from DDoS attacks exceed the costs of cache capacities, ICPs will utilize a desirable amount of cache. A CDN, in contrast, consists of a global cache infrastructure, a set of distributed redirectors, a network operating center, and third-party audits. It is in the owner's best interest to oversell the capacity and maximize profits. From one ICP's point of view, although Service level agreements (SLAs) are specified in terms of long-term contracts, the operation of CDNs is not directly committed to its payoffs since the redirectors (i.e. the allocation of capacity) are designed to optimize overall performance of the whole CDN. There are potential conflicts among users of a CDN. Therefore, the best strategy for each individual ICP is often to over-utilize the caches, and the incentive chain is still broken in the CDN environment.

This example of CPNs illustrates that cache trading shifts incentive for ICPs to ISPs. A dynamic pricing mechanism based on network conditions and demand can potentially lead to more efficient allocation of cache capacities. However, the CPN itself does not provide information for resource discovery and evaluation. In section 5, we will discuss network mapping, which is a service that facilitates dynamic resource allocation.

Cost Assessment

A solution to DDoS attacks would have little practical value if the costs involved in implementation outrun the benefits. The costs include one-time investment (acquiring additional devices and redundant resources), operation costs (filtering, monitoring, collecting and exchanging information, etc.), and administrative costs (configuring, billing, auditing, and disputing).

Under the current Internet architecture, the cost of CDNs and server-side detecting/filtering is very high because the huge investments in monitoring infrastructures and redundant resources drive prices up. For example, Akamai invested millions of dollars in building their caching network, measuring networks, and operating center, which only support a small number of customers. For server-side filtering, some products provide moderate control to enhance performance without a significant measurement network. Since it is difficult to collect information about traffic (such as usage patterns and the fingerprints of attacks) and perform filtering at edge points, the solutions require very complicated algorithms and processing capacity, which result in expansive hardwire implementations. For example, with prices of intelligent routing solutions as high as $250,000, customers can choose the quality and cost of services from multiple ISPs. For the DDoS attack-detection and mitigation, the price is as high as $32,000 for filter/IDS based products (According to the review of Network World, Sep 2, 2002).

A large up-front cost to customers limits the potential customer base. Since DDoS attacks are rare events for a specific firm, considering a six- to nine-month return on investment (ROI), the value of these solutions is very small. The overhead of extra layer and filtering may degenerate the performance of networks and yield a negative ROI. Moreover, a management tool that combines traffic monitoring and filtering could cost more intensive networking jobs. To drill down and fine-tune a network, more labor hours

are required in administration and operation tasks, such as the analyses of raw data and the configuration of routing or filtering.

Compared to CDNs and the server-side solutions, direct investment in bandwidth and caching is more cost-effective. The report by TeleGeography indicates that costs of IP transit connections in Europe and the U.S. have fallen continually over the past three years. For instance, the price of STM-1 (155 Mbit/s) links is getting as low as $50 per Mbit/s (based on a report from Band-X Ltd., a bandwidth exchange). Moreover, ISP peering can largely reduce the cost of bandwidth. For example, in a promotion, the Equinix San Jose Internet Business Exchange (IBX) offers a DS3 (45Mbps) transport circuit for $1000/month (Norton 2002).

In current practices in the ISP industry, since almost all pricing schemes in bandwidth and caching are flat-fee based (even in the transit-based traffic exchange, a monthly fee is charged based upon a peak rate traffic sample), firms choose to acquire extra capacities instead of investing in pro-active solutions for potential DDoS attacks. By utilizing the existing resources in networks, cooperative solutions, such as CPNs, reduce the requirements of traffic measuring and analyzing, as well as corresponding administrative costs.

OTHER TECHNOLOGIES IN IMPLEMENTING THE REMEDIES

When it comes to the implementation of potential remedies, some emerging technologies come to our attention since they have similar characteristics of usage-based pricing and CPN, and significantly affect the effectiveness of these remedies. In this section we focus on two technologies, overlay networks and Internet mapping, and give a brief review of others.

Overlay Networks

Internet users are often charged not only by their ISPs. For example, a music fan may need to pay fees to entertainment websites or commercial Peer-to-Peer (P2P) networks for access to online albums. In cases where users/regional ISPs pay multiple fees in which the fee charged by their direct regional/backbone ISPs is only a small portion of the total payment, an effective incentive structure against DDoS attacks requires that the total fee has to be incremental as we discussed before.

Beside the payment to the directly connected ISP, each of the fees an Internet user pays allows her to utilize a specific Internet service. Often, a specific Internet service can be viewed as an application-level network that is based on the general infrastructure of the Internet, referred to as an overlay network. In an example of music sharing, an overlay network can be a content intermediary provided by an entertainment website, or simply a distributed P2P network. Overlay networks for various purposes have been heavily studied in recent years by computer scientists (see, for example, SkipNet Harvey et al. 2003), and are arguably viewed as the future organizational form of a scalable Internet computing environment. Ideally, multiple overlay networks are expected to share the common Internet infrastructure for their computing and communication needs, while at the same time maintain their relative independence in performing their respective tasks.

In the scenario of overlay networks, the total payment of an Internet user often consists of two parts: one part is the payment to the direct ISP for Internet access – in other words, payment for communication infrastructures; the other consists of a collection of payments, each of which goes to a specific overlay network.

Fee structures can vary a great deal across different overlay networks. While accessing up-to-date stock quotes in a financial network is often expensive, sharing

family pictures and videos in peer-to-peer networks can be cheap or even costless. Not surprisingly, according to the usual economic principles of supply-demand and competition, the value of services and the switching cost for users to choose other similar alternatives determine the fee structure of an overlay network.

Nevertheless, the discrepancy in fee structures among various overlay networks, and the independent nature of all overlay networks, may thwart the effective defense against DDoS attacks if the total fee a user/regional ISP pays lacks incremental merits. For example, even if all other overlay networks as well as ISPs charge usage-based fees, a single overlay network (think of Napster or Gnutella) that does not implement an incremental fee structure will significantly flatten the total payment all users paid, which in turn weakens the incentive chain against DDoS attacks.

The situation is further worsened by the possibility of free riding. Recall all overlay networks share the same Internet infrastructure. Users' strong demand for some highly valuable overlay networks (e.g. financial overlay networks) may provide strong incentive for ISPs to heavily invest in bandwidth. Knowing this, other overlay networks may take advantage of the large bandwidth for other low-valuation but high traffic purposes. One extreme example is frequent network congestion in university networks caused by intensive music sharing activities.

The key to solving the challenge that overlay networks pose for implementing a usage-based fee is incentive alignment among various overlay networks: in our view, while maintaining functional independence, the cost structure of each overlay network should be tied to the activities of other overlay networks since the increasing traffic in one network negatively affects all others as they all share the common Internet infrastructure. Although specific measures for incentive alignment depend on specific overlay networks, which remain to be studied, the general principles can be found in

existing economic literature on the trading of pollution rights (Ledyard and Szakaly-Moore 1994). Intuitively, users in an overlay network should pay not only the direct costs of using that overlay network, but also the damage to other overlay networks. Another useful economic tool is the comparative statics (Currier 2000), which helps to adjust incentives on a micro-level in complex situations.

Internet Mapping for Optimal Supply Chain Selection

Another related technology is Internet mapping. Note that for any content request, the choice of suppliers may not be unique: a client can request a certain chunk of content from several candidate content servers; similarly, an ICP can have its content cached at selected cache servers out of a large pool of candidates. Such a selection problem can be solved if the client or the ICP in question has a complete view of Internet topology, and has up-to-date information on related delay information. Nevertheless, maintaining a complete map of Internet topology and corresponding information requires a considerable amount of storage and processing capability from each computer.

An alternative approach to optimizing supply chain selection is Internet mapping, where the Internet is hypothesized to be a multi-dimensional Euclidean space in which all network devices reside (Ng and Zhang 2002). Giving a set of representative landmark nodes, all other devices on the Internet measure the delay from them to the landmark nodes, and accordingly calculate their locations in the Euclidean space. Based on their locations, any two devices on the Internet can approximately estimate the delay between them by simply calculating the distance of the two positions.

Based on Internet mapping, it is straightforward to see that effective traffic filtering and content caching depend on accurate mapping information. Empirical studies show that Internet mapping is able to produce good approximations to the actual delay between any two devices on the Internet (Ng and Zhang 2002), if all devices truthfully

report their locations and delay information. However, this is another big "if", as Internet mapping could go wrong even if we have good algorithms for it.

In Internet mapping, a network node calculates its position based on the positions of several reference nodes and the delay to each of them. The assumption in this process is that all reference nodes will truthfully report their positions. However, this is not always true as reporting the truth may not be in the best interest of one or more reference nodes. As an example, consider a cache server that resides at a high-bandwidth neighborhood so when its neighbor ISPs are in high demand, it is often in the best position to help. Knowing this, the cache server will expect a large amount of requests from its neighbors if they are affected by DDoS attacks. If the ISP owning the cache server is not compensated enough for helping cooperative caching, or if the ISP simply does not want to participate in cooperative caching, it can avoid the cooperation by cheating and responding with an erroneous and remote position to all requests. In this way, it avoids heavy traffic at the local site but also reduces the effectiveness of cooperative caching.

The incentive issue in Internet mapping is difficult to solve since a successful solution needs to give all participating reference nodes enough incentive to truthfully report. When deciding whether to report its true location or not, a network node needs to consider all activities it is currently participating in, which could be highly dynamic. A useful reference on incentive alignment among large numbers of network nodes is Stahl and Whinston (1994). In the scheme of defending DDoS attacks, network mapping is one of essential overlay technologies, which provides applications with information and extra control beyond low-level networks. Therefore, end users, such as ICPs, can implement flexible defending solutions with the principle of end-to-end arguments (Saltzer, Reed, and Clark 1984).

Key Players in Reconstruction of the Internet Infrastructure

DDoS attacks are an extreme case of Quality of Service (QoS) problems on the Internet. Because of the growing concern about the reliability of the Internet, many companies have addressed QoS issues from various aspects. Some of them have embedded resource-trading markets and have implemented usage-based fee structures. The following table summarizes the major players and commercial technologies employed in the reconstruction of the Internet infrastructure.

Category	Technologies	Leader companies
ISPs	ISP peering, caching, edge control	Cable & Wireless, GTE, PSInet, Sprint, Uunet, AT&T, Qwest ; Cox, etc.
Border Route Services	Intelligent routing, Multi-homing, Traffic Classification	Internap (Netvmg and Sockeye), Proficient Networks, RouteControl; P-Cube etc.
Content Delivery	Content Distribution Network (CDN), Global Load Balancing (GLB)	Akamai, Speedera; Cable & Wireless, NTT; Coyote Point Systems, F5 Networks.
Network Monitoring	Active Probing, Performance Monitoring, Traffic Engineering	Keynote, F5 Networks, Mercury Interactive Corp, Gomez.

Table 1.1. Major players in reconstruction of the Internet infrastructure

Given the ability to managing Internet traffic directly, ISPs have the biggest impact on the structure of the Internet. ISP peering and caching are widely used to improve quality of service. However, peering and caching have a limited influence on Internet architecture since only the owners of resources can benefit from these technologies. In other words, the usage of peer connections and cache capacity are not transferable. Therefore applications cannot get direct support from ISP peering and caching. As a result, Internet topology becomes extremely complicated and a large amount of resources are wasted.

To solve this problem, several companies provide additional resources directly to content and service providers. Extra resources, such as alternative Internet connections and storage, can be acquired by particular applications and increase the reliability of the services. Intelligent routing is an edge point solution that directs outbound traffic among multiple Internet access connections applies according to the characteristics of differnet communication channels. After acquiring Netvmg and Sockeye, Internap has become the leader in providing intelligent routing solutions. By building an overlay network that continually analyzes the traffic situation on major Internet backbones, Internap selects the path of least resistance for its clients to deliver content faster and more reliably. When congestion occurs at one of the connections, a multi-homing ICP can still publish its content through other connections. However, intelligent routing only partially mitigates DDoS attacks. Since only outbound traffic is under control, a DDoS attack will block all incoming requests for services from one or more connections and therefore shutdown the services in a particular region.

In content delivery solutions, the Content Distribution Network (CDN) and Global Load Balancing (GLB) impose controls on inbound traffic. Redirectors (such as dynamic DNS servers) distribute incoming requests among a set of cache servers or hosts using algorithms with the perception of network delay and the performance of servers.

It is clear that each player has strengths and weaknesses. Cross-sector collaborations are forming among the players: ISPs start to adopt tiered-pricing systems based on traffic classification; companies in different industries share resources to reduce sunk-costs; the intelligent routing industry has been consolidated to provide cost-effective infrastructures. As an overlay service, a network monitoring service is a perfect glue to facilitate the collaborations in defending against DDoS attacks.

Using software agents deployed over the Internet, network monitoring services produce information on Internet topology, traffic reports and tracemaps, and help IT personnel quantify bandwidth utilization and delay. This will help them assess the possibilities of overload and the need (or lack of need) to purchase additional capacity. For example, as a pioneer of network mapping, Matrix NetSystems (acquired by Keynote Systems in 2003) has built a global monitoring infrastructure with the help of ISPs to analyze Internet performance. Through the collaboration with a large set of backbone ISPs (e.g. PSInet) and hosting companies (e.g. Internap), Matrix NetSystems deployed measurement computers, called "beacons", in different geographical locations worldwide. The beacons use active probing technologies, such ICMP pings, to measure Internet IP transit delay and costumer-specified parameters. With the "beacon network", Matrix NetSystems could measure specified routes precisely and became one of the three major venders of DDoS alert services for the Department of Homeland Security.

As a strategic solution, organizations can use this diagnostic information to determine which physical link (ISPs, nodes, etc.) in the Internet infrastructure is responsible for performance degradation and, in turn, remove the bottlenecks. Furthermore, real-time traffic information can be used in tactic solutions, such as directing traffic in intelligent routing and trading bandwidth and cache capacities. Network mapping services facilitate information collection and payment exchange to achieve an integrated defense scheme.

CONCLUSION

The major message we want to convey in this paper is that, in dealing with DDoS attacks, industry and academia have long ignored the incentive aspect of the problem, which turns out to be the key in defeating DDoS attacks. We then argue that two incentive mechanisms, the usage-based pricing and CPN, can effectively support

effective technological solutions, such as cooperative filtering and cooperative caching. In practice, we expect to see a combination of both incentive mechanisms and the two technological solutions: blocking a flood is quite effective, but sometimes diluting and thus smoothing the flood may be more cost-effective.

As an important note to practitioners, the incentive mechanisms and technological solutions in this paper apply to but are not exclusive to DDoS attacks. Unless a DDoS attack is highly likely to happen and the consequence is significant, no business will spend a huge amount on new incentive mechanisms or technological solutions solely for DDoS attacks. Rather, new solutions must address a much broader range of QoS problems, in which the DDoS attack is just one extreme case that one needs to be aware of. As an example, ISPs may initiate usage-based pricing mainly for better QoS in transmitting legitimate but prioritized traffic – as long as it has the incremental payment characteristic we discussed in this paper, it also helps against DDoS attacks.

CHAPTER 2: Defending Wireless Infrastructure Against the Challenge of DDoS Attacks

This paper addresses possible Distributed Denial-of-Service (DDoS) attacks toward the wireless Internet including the Wireless Extended Internet, the Wireless Portal Network, and the Wireless Ad Hoc network. We propose a conceptual model for defending against DDoS attacks on the wireless Internet, which incorporates both cooperative technological solutions and economic incentive mechanisms built on usage-based fees. Cost-effectiveness is also addressed through an illustrative implementation scheme using Policy Based Networking (PBN). By investigating both technological and economic difficulties in defense of DDoS attacks which have plagued the wired Internet, our aim here is to foster further development of wireless Internet infrastructure as a more secure and efficient platform for mobile commerce.

INTRODUCTION

The wireless Internet has become an exciting realm for m-commerce at an amazing speed. The estimated number of wireless subscribers was 109 million in December 2000 in the US alone, according to a semi-annual wireless industry survey conducted by Cellular Telecommunications Industry Association (CTIA 2000). It represented an increase of 27.2% from a year earlier, adding nearly 23.43 million new users. According to a new study released by Strategy Analytics, the global cellular market will grow at an annual rate of 17% over the next five years, reaching $700 billion with 1.4 billion global wireless subscribers by 2005 (Strategy Analytics).

M-commerce is not a simple duplication of e-commerce upon wireless devices. As pointed out by market research institutions including Goldman Sachs (2000) and Bear Stearns (2001), "m-commerce is about information and transactions that are timely."

Is wireless infrastructure ready for time-sensitive m-commerce? From a technological perspective, it is ready for anytime, anywhere access. 3G wireless technology also enables high-speed access. However from a security perspective, time-sensitive m-commerce is vulnerable to network delays or even network denial caused by a dangerous type of security problem – the Distributed Denial-of-Service (DDoS) attack – that has been much publicized but seldom understood completely (Geng and Whinston 2000, Internet Security Systems).

Due to the time-sensitive nature of m-commerce, it is not surprising for wireless infrastructure providers to carefully plan the radio spectrum allocation and pricing to avoid any predictable congestion. Given the huge cost of radio spectrum rights, they also have enough incentive to defend against most security risks through constant and prompt patching of system security holes and real-time monitoring. These remedies, however, target unauthorized intrusions. A DDoS attack, on the other hand, never tries to break into the victim's system. On the wired Internet, attacks against well-known sites (Fisher and Callaghan 2001, Haney 2001, Hopper 2000) have repeatedly proved the lack of an effective defense. As Geng and Whinston (2000) pointed out, effective defense is unlikely to appear on the present wired Internet as there lacks an incentive structure to push cooperation on the wired Internet.

DDoS attacks are not a serious problem to the current wireless Internet, in part because of the extremely limited and often non-programmable functionalities of current mobile devices. However, our research strongly suggests that DDoS attacks can be a real threat in the near future given the increasing computational power, network bandwidth,

and users in the wireless Internet economy. Two significant events have already occurred. First, in the summer of 2000, there appeared the first preliminary virus against mobile phones (Dennis 2000). Furthermore, Eugene Kaspersky, head of anti-virus research at Kaspersky Lab, a Moscow-based anti-virus company, once commented on this virus:

"This is not the first and obviously not the last security breach discovered in mobile phones. Moreover, I believe as more functionality is added to mobile phones, it will result in more breaches being found."

The second event was the emergence of the first DDoS attack tool toward mobile phones, known as the SMS flooder (Sherriff 2000). It tries to use the wired Internet to attack a wireless victim. First it proliferates through Microsoft Outlook just as the Melissa virus (see http://www.cert.org/advisories/CA-1999-04.html for details) does. Then it commands all infected Microsoft Outlook software to send short messages (SMS-messages) to a certain victim's mobile phone to inundate it. The potential hazard is not only in the blocking of communications but also in the high financial cost if pricing is usage-based.

The two events mentioned above show that the DDoS attack directed towards the wireless Internet is not only a theoretical possibility, but also a real and evolving threat. However, research is lacking as to what forms DDoS attacks against the wireless Internet will possibly take and how they can be defended effectively – technologically, economically and in terms of cost. This article tries to answer these two questions. We start by briefly reviewing the mechanism of DDoS attacks in section 2.

In section 3, we analyze new features of the wireless Internet infrastructure and possible DDoS attack forms. Since various standards for the wireless Internet are still emerging, we discuss three infrastructure schemes – the Wireless Extended Internet, the Wireless Portal Network, and the Wireless Ad Hoc Network. Intuitively, possible forms

of DDoS attacks include not only ones that are found on the wired Internet – e.g., attacking e-business servers – but also new forms such as attacking the radio spectrum that is naturally a scarce resource. Another new attack form is the attack across both the wireless and wired Internet. Given the differences in computational power and the bandwidth between wired and wireless devices, it is easier for an attacker to use wired devices to initiate cross platform attacks toward wireless devices.

Section 4 proposes a conceptual model for defending against wireless DDoS attacks. In this model, we address three issues. First, we consider technological solutions based on the analysis of possible attacks. Secondly, we evaluate economic costs and benefits involved in motivating the usage of these technological solutions. As the attacks in February 2000 have shown, the biggest barrier in defending against DDoS attacks is the lack of economic incentives for Internet users to cooperate (Geng and Whinston 2000). The third is the implementation issue – i.e., how to construct both technological solutions and incentive structures in a cost-effective way. Section 5 concludes this article.

MECHANISM OF DDOS ATTACKS

The DDoS attack is the most advanced form of Denial-of-Service (DoS) attacks. As the name suggests, the DDoS attack is distinguished from other DoS attacks by its ability to deploy its weapons in a "distributed" way over the Internet and to aggregate these forces to create lethal traffic. What drives hackers to move DoS attack tools to the distributed level is the ever-increasing security in potential victims' systems in this cat-and-mouse game. Figure 2.1 outlines the evolution of both attacks and defenses. For a detailed explanation see (Geng and Whinston 2000).

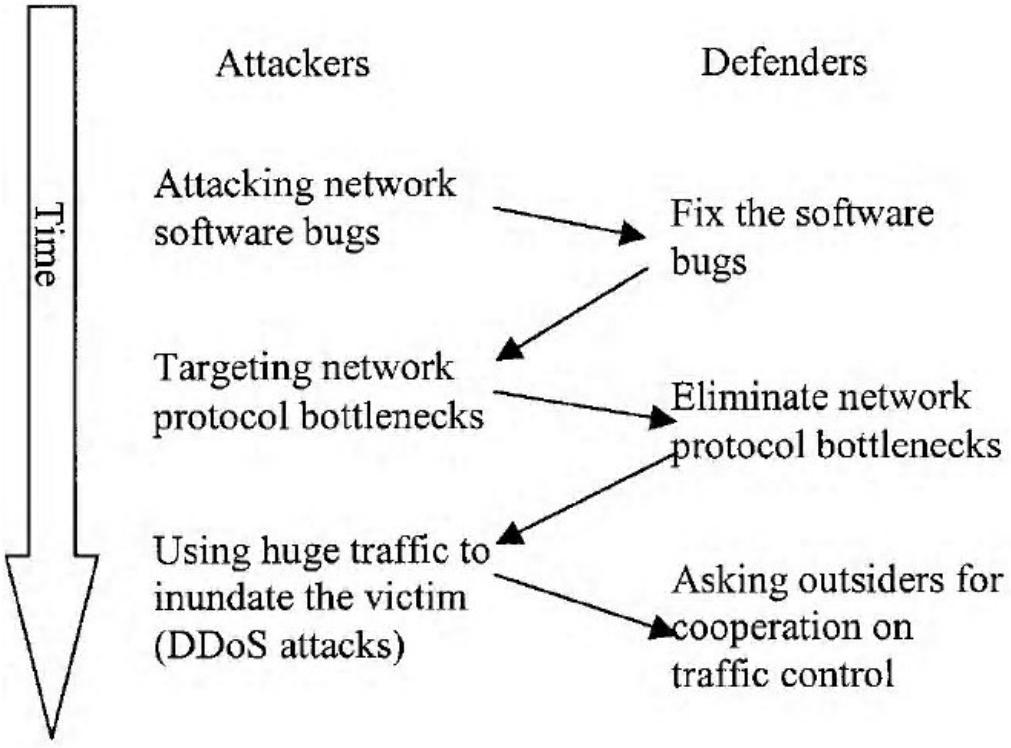

Figure 2.1. The evolution of attacks and defenses in DoS attacks.

Although the presence of bugs in network software makes the most primitive DoS attacks still viable, e-businesses are more sensitive and prompt than before in protecting their system security by using intrusion detection software and by applying patches. As a result, the most frequent and harmful DoS attacks are in distributed form. DDoS attacks are distinct from all prior DoS attacks in that they never try to break into the victim's system, thus making any security defense irrelevant. There are numerous variances of DDoS attack tools, all of which share a similar structure.

A typical DDoS attack structure is shown in Figure 2.2. The attacker first gets control of several master computers by hacking into them. Then the master computers further get control of more daemon computers (also called zombie computers), often by using some automatic intrusion software. Such a hierarchical structure is difficult to trace

back. Finally, a command from the attacker can synchronize all daemons to send junk traffic to the victim, often a well-known site in e-commerce, to effectively jam its entrance and block access by legitimate users.

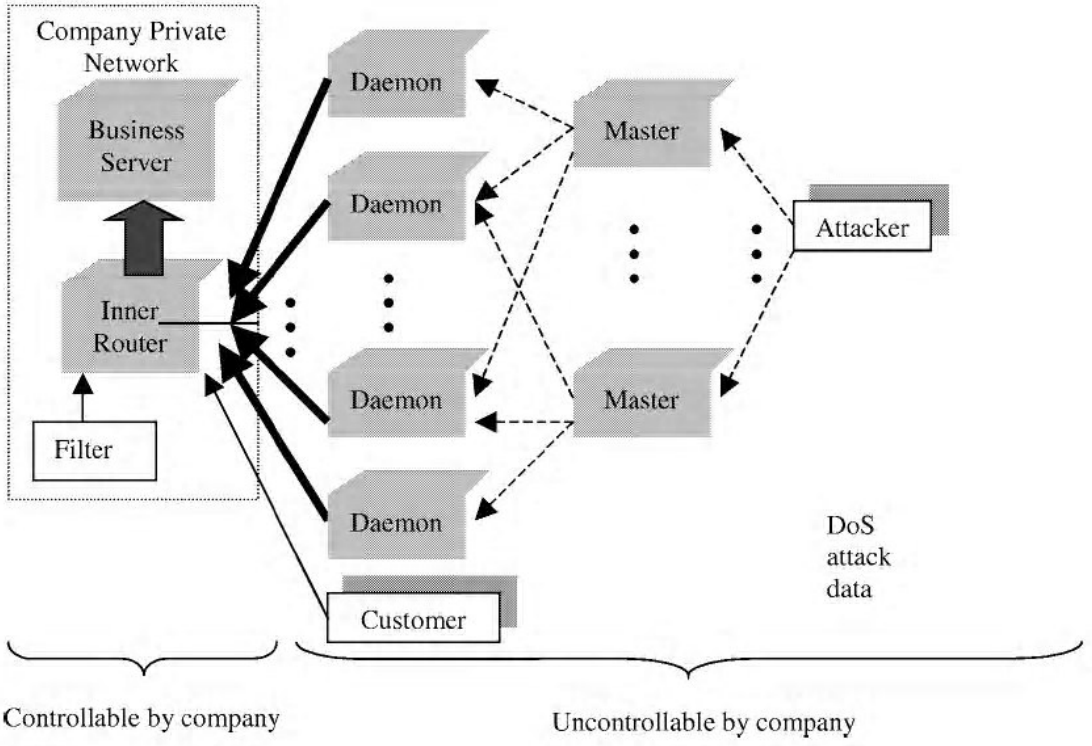

Figure 2.2. A typical DDoS attack structure.

In practice, various DDoS tools differ in terms of the hierarchical structure, attacking packets generated, corresponding attacking targets, and the encryption of communication. For a more comprehensive list and analysis, see Packet Storm at http://packetstorm.securify.com/ and David Dittrich's articles at http://staff.washington.edu/dittrich/misc/ddos/. It is worth noting that all these DDoS attack tools are available in source codes on the Internet and new versions keep emerging. New and "improved" versions are more complicated in the way they conceal attacking traffic and in encryption methods, making the defense more difficult.

For the wired Internet, Geng and Whinston (2000) show that three problems lead to the proliferation of DDoS attacks: the insecure Internet, a lack of an effective way to control junk traffic, and IP spoofing.

INFRASTRUCTURES OF WIRELESS INTERNET AND DDOS ATTACKS

Two aspects differentiate the wireless Internet from the wired Internet. From a technological perspective, differences between wired and wireless networks are due to link characteristics and user mobility (Naghshineh et al. 1996). Compared to coaxial cable, DSL, and fiber, the wireless link is characterized by high cost, volatility, high error rates and relatively small transmission capacity. Because of shared radio spectrum, communication can be interfered by competing users, other equipment, evil intent hackers, or even natural phenomena. In terms of user mobility, the user–network interface (UNI) in a wireless environment keeps changing throughout the duration of a connection.

From an economic perspective, the wireless infrastructure is likely to be an oligopolistic market, while the wired infrastructure is open to competition. The wireless infrastructure market is dominated by a few cellular phone carriers and wireless equipment providers with different communication standards and private technologies. In addition, the high cost of radio spectrum licenses and geographic constraints make an entry to the wireless access market difficult.

Based on different application models, the wireless Internet can be categorized into three different infrastructures: the Wireless Extended Internet, the Wireless Portal Network and the Wireless Ad Hoc Network. The Wireless Extended Internet is merely an extension of the wired Internet for mobility convenience. Wireless Portal Networks are developed and privately owned by wireless telecommunication providers, thus are highly

centralized. Unlike the former two, Wireless Ad Hoc Networks have no client-server structure.

Wireless Extended Internet

In theWireless Extended Internet, wireless technology is used only for the last mile. Wireless access providers, or wireless ISPs, connect mobile devices to fixed networks via radio frequency (RF) channels. The traditional Client/Server architecture, as well as existing transport layer protocols (usually TCP), is also used for the Wireless Extended Internet. Therefore, DDoS attacks seen in the wired Internet are still feasible in the Wireless Extended Internet.

Attacking devices using aggregated traffic

Tens of millions of cellular phones, laptops and palmtops are expected to use wireless connections to access the Internet in the near future. Although transmission rates in wireless networks are much lower than those in wired networks, potential DDoS attacks are still feasible if large population of mobile units are involved. Thus, wireless data packet traffic is a potential avenue for DDoS attacks.

Attacking the asymmetric structure

Mobile devices have less computation and communication capabilities than those of fixed devices. A DDoS attack, even launched by a small number of powerful fixed computers, can effortlessly disable a large range of mobile devices. Wireless Internet content servers – such as WAP, wireless game, and mail (instant message) servers – are often optimized for small throughput and timely response. Thus, they are especially vulnerable to DDoS attacks compared with traditional wired servers.

Furthermore, there may emerge new forms of DDoS attacks taking advantage of new characteristics of the wireless communication.

Attacking the radio spectrum

The limited availability of radio spectrum is always the bottleneck in a wireless network. Even if license-free RF bands (such as the ISM band in the US) are used and micro-cell and pico-cell technologies are employed to expand transmission rates, it is still a scarce resource as the number of users and the demand for bandwidth increase. Technological research on wireless bandwidth allocation and admission control relies on stochastic theories, assuming that users will not use their devices all at the same time. Therefore, the total communication bandwidth can be far less than the total communication capacity of all wireless devices. However, a DDoS attack deliberately coordinates wireless devices to send out synchronized traffic, which can easily consume all spectrum resources or at least significantly reduce the capacity of communication channels for normal traffic.

Avoiding tracing back by mobility

The IETF Mobile IP protocol is a significant step towards enabling nomadic Internet users. Most research on security in Mobile IP deals with registration, authentication, key management and encryption. However, Mobile IP still has flaws that DDoS attackers can use in addition to conventional security problems. For example, the Mobile IP protocol requires two IP addresses: the home address and the care-of address. The home address is permanently assigned to a mobile device, while the care-of address is temporarily assigned by the visiting foreign network. Similar to IP-spoofing, the Mobile IP protocol allows a mobile device to send out IP datagrams using its fixed home address even if it roams away. Some extensions of Mobile IP are also sources of concern. For example, the Non-Disclosure Method (NDM) prevents the tracking of user movements by third parties and gives mobile users control over the revelation of their

location information, according to their personal security demands (Fasbender et al. 1996). As a result, victim sites will find it difficult to trace sources of DDoS attacks.

Wireless Portal Network

Learning from America Online's success, most wireless operators are using various "walled garden" and partnership approaches. Since they own coveted spectrum licenses and cellular phone user bases, these operators have strong bargaining power over all their business partners. Therefore, they are in a better position to secure additional revenue streams, including slotting fees for portal placement, a slice of m-commerce revenues, and fees from location-based services. Such an extension of their business will transform them into wireless portals (see Figure 2.3). The most cited example is NTT DoCoMo, for which 5.9 million users signed up with its i-mode service during the last four months of 2000.

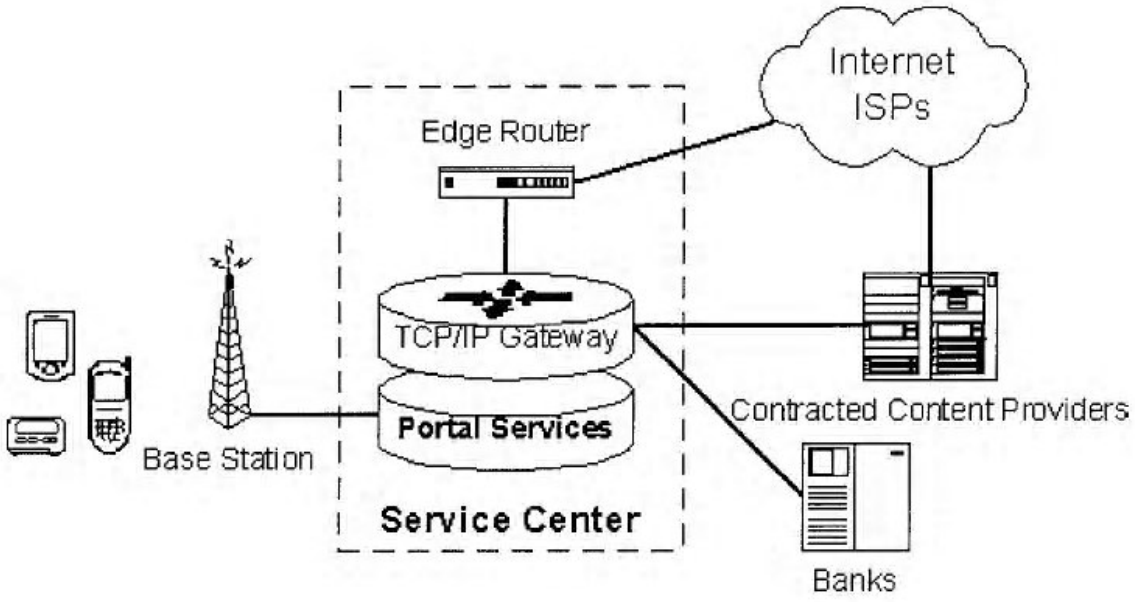

Figure 2.3. The architecture of the Wireless Portal Network

The Wireless Portal Network is based on the typical Client/Server architecture. Mobile clients (usually cellular phones, smart phones, and specific PDAs) embedded

with compact Operating Systems communicate with base stations through wireless packet-switched data networks. All requests are passed to the service center through the telephone network and signaling systems. Similar to the Service Control Point (SCP) in an Intelligent Network, the service center keeps user information and provides portal services, contracted services, and public Internet services. Portal services are kernel services in a Wireless Portal Network, which maintain user profiles and billing databases and provide location-based service and other real-time services. Application requests and responses will not be encapsulated in IP packets. Thus, they have the lowest latency. For contracted services, the requests are translated into TCP/IP protocol streams by the TCP/IP gateway and served by contracted content providers. Dedicated lines and reserved paths guarantee security and QoS. For public Internet services, Internet access requests will be passed from edge routers to the backbone.

Clients, contracted content providers, and the service center become a walled community, i.e., a reliable "security island". This architecture is more secure than the Wireless Extended Internet because a Portal Network screens all clients and most servers located in the public Internet. It is difficult to launch attacks from outside the island. However, with increasingly powerful phones, such as Java phones that could be infected with DDoS zombie viruses, the network could be vulnerable to internal attacks.

Attacking the radio spectrum

Because Wireless Portal Networks primarily employ existing cellular phone systems (single-hop), a base station is the only entry to a specified cell. In major cities and crowded airports, it is common to have calls dropped in mid-sentence. Sometimes making a connection is impossible. Mimicking this natural congestion, it is possible to disable a particular base station – e.g., the one serving an important conference nearby – by simultaneously sending connection requests and a mass of traffic from mobile

zombies. As a result, all wireless devices within this cell will not be able to connect to the network. In some cases, even control channels can be blocked. In a Personal Communications Services (PCS) Network, when a Visitor Location Register (VLR) fails and broadcasts a re-registration request to all Mobile Stations (MSs), registration messages sent by MSs will cause a natural traffic jam (and thus, collisions) in the reverse Digital Control Channel (DCCH) (Haas and Lin 2000). Therefore, if the MSs have more control over the DCCH, they can block the channel and make VLR busy with recognizing fake identities. Then the traffic channel will be of no use even if it is available.

Attacking TCP/IP gateway

The TCP/IP gateway translates between wireless bearer protocols and the Internet TCP/IP protocols. It is one crucial bottleneck in the Wireless Portal Network. Abundant computing capability and enough links are extremely important for it to provide a security protection for mobile terminals and inner servers against attacks from the public Internet. If one has to shut down the gateway, the Wireless Portal Network will be isolated from the public Internet and make all outside services unavailable.

Attacking value-added services

It is difficult to attack value-added services since dedicated lines will be used for such crucial services as banking and trading and some content servers are embedded into portal services, like location services. All these services are invisible outside the portal networks and will survive under outside DDoS flooding. However, there might be sophisticated methods to launch attacks from devices within the portal network.

Wireless Ad Hoc Network

A Wireless Ad Hoc Network (also called multihop network or Peer-to-peer wireless network) is formed temporarily by a group of mobile devices, which have a common mission or interest. Adhering to a strict admission policy and communication rules, all these devices form a special community of equals to share information. There is no designated client or server. All members communicate over wireless channels directly without any fixed networking infrastructure or centralized administration. In this structure, all mobile hosts communicate with each other in a wireless multi-hop routing style. Each mobile node maintains all the links within the defined radius (called zone) and acts as a router in the network. If a member is out of its destination member's zone or it is not in a line-of-sight, all messages between them must pass through one or more routers. All members are free to move around and join and leave a network at will without any technical difficulties, subject to admission control. The routing scheme is adjusted dynamically according to the changing network topology.

Analogous to the Internet that evolved from the simple DARPA net, the Wireless Ad Hoc Network has the potential to grow into a World Wide Wireless interconnected network. Wireless Ad Hoc Networks were first recognized as an important issue in the military communications arena in the 70's. Several systems have been deployed for the Tactical Data Systems, such as Link-16 in the US Navy Airborne and Shipboard systems. Following the wide deployment of mature wireless technologies, the Wireless Ad Hoc Network is receiving more attention for commercial applications, such as team collaboration applications, networking intelligent sensors and cooperative robots, etc.

The Ad Hoc Network is the best architecture against DDoS attacks. First and foremost, it has no central server. Secondly, it may implement strict admission policies making it very hard for outsiders to hack into the communication infrastructure. Multi-

hopping reduces transmitter power and protects network capacity via spatial reuse. Because there is no central point and no crucial resource, any blocked route can be substituted by redundant links. In addition, the community can reject an abnormal member by voting based on certain admission policies. Dynamic routing protocols and mobility of the network components give Ad Hoc Networks a self-adjusting capability under attacks.

It is unlikely that the Wireless Ad Hoc Network will be restricted to a small geographical region. Hybrid architecture could be used to expand the range of such networks. Members can communicate with one another via the local RF network within a regional wireless community, and with other members located anywhere within reach of the commercial telephone system through wired relay services. With the help of the dual-membership hosts, interconnecting different communities will result in the World Wide Wireless network. Wireless communities can also be attached to conventional fixed data networks to expand application possibilities. For instance, home-networked appliances based on Bluetooth technology can be remotely controlled through the Internet. For military use, a complete networking system, called the AEGIS Broadcast Network, has been implemented for tactical data systems in the US Navy. It connects, monitors, and controls all military units on both coasts, the Gulf of Mexico, Japan, etc. The interconnection among Wireless Ad Hoc Networks through wired relay services creates a complex network topology, in which critical points can be attacked. First, attacks against dual-membership hosts may effectively disable the interconnections among different Ad Hoc Networks. Secondly, directory services, which are indispensable for large scale interconnected Ad Hoc Networks, are also possible targets for DDoS attacks. This is similar to the case in the Internet where DNS servers and catalog servers are frequent targets of DDoS attacks. In a word, the World WideWireless network could be subject to

all forms of DDoS attacks that exist on the Internet if it evolves towards an asymmetric infrastructure.

DEFENDING AGAINST DDOS ATTACKS ON THE WIRELESS INTERNET

In the event of a typical DDoS attack, the victim alone cannot effectively defend herself/himself. Cooperation among all involved parties is indispensable. Figure 2.4 presents our conceptual model for defending against a DDoS attack, which illustrates a two-layer coordinated defense problem and an implementation problem.

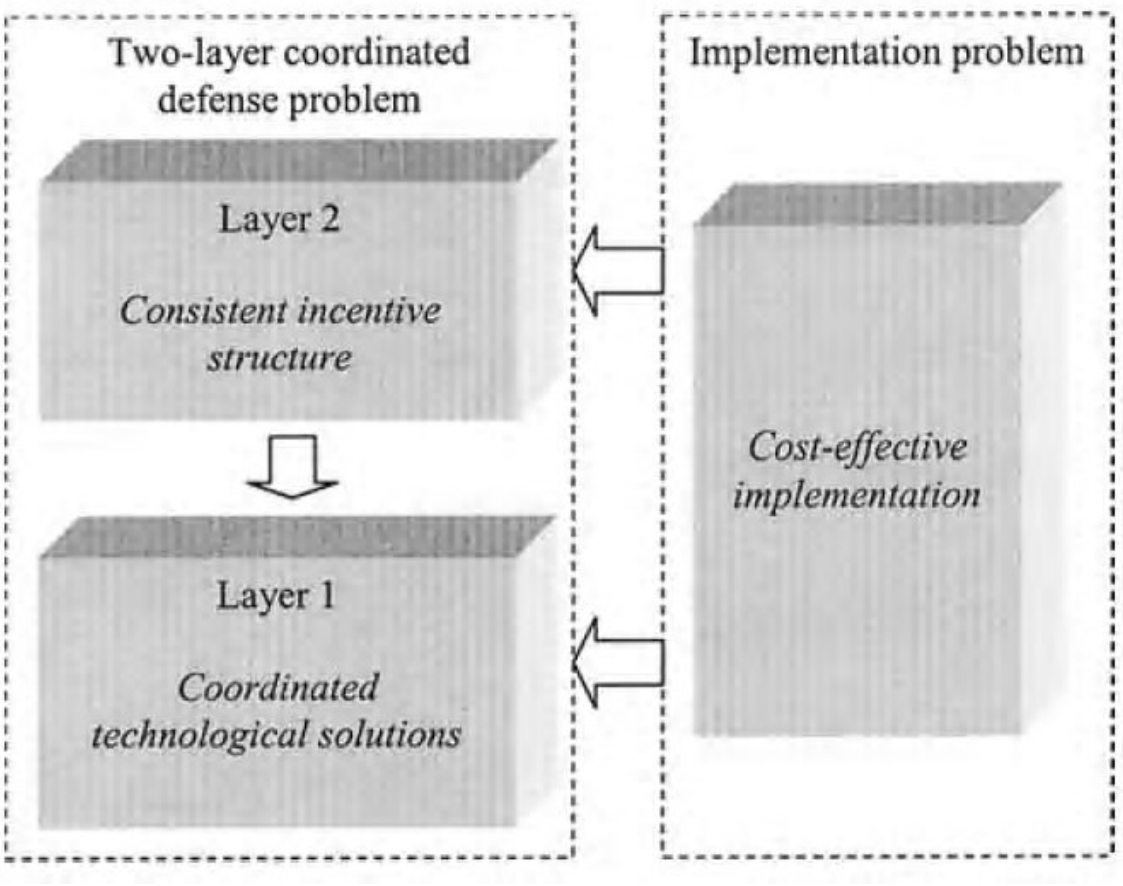

Figure 2.4. The conceptual model for defending against the DDoS attack

In the two-layer coordinated defense problem, the first layer focuses on effective coordinated technological solutions. The second layer deals with the incentive mechanism that, in an economic perspective, makes people involved in a DDoS attack

feel that cooperating with each other is the best strategy. In past practice, unfortunately, little attention has been paid to this second layer problem compared with the public focus on technologies. Ironically, this incentive problem causes the most headaches in practice (Geng and Whinston 2000). As a solution, we propose to use usage-based fees as the foundation of the incentive mechanism.

The objective of the implementation problem is cost effectiveness, which arises as a crucial problem because defending against DDoS attacks may require an overhaul of the current network infrastructure. For instance, the implementation of a usage-based fee scheme on the wireless Internet – as well as on the wired Internet if we consider the cross-border attacks between the wired and wireless Internet – has strong demands on the network's ability to audit and manage traffic. As an illustrative example, we present an implementation scheme based on the Policy Based Networking (PBN) framework.

Coordinated technological solutions

There are four types of coordinated technological solutions, as shown in Figure 2.5.

Two comments are necessary for Figure 2.5. First, different solutions can coexist to achieve a better defense. For example, user-level traffic control and coordinated filters can be implemented simultaneously to be more effective. Second, as in the wired Internet example, coordination is often required to be global, whereas in the wireless Internet case local coordination may suffice. For example, to avoid an attack on radio frequencies in a certain geographical area, it is sufficient to require coordination only among involved wireless devices and base stations in that area. Below we analyze the characteristics of these four coordinated technological solutions.

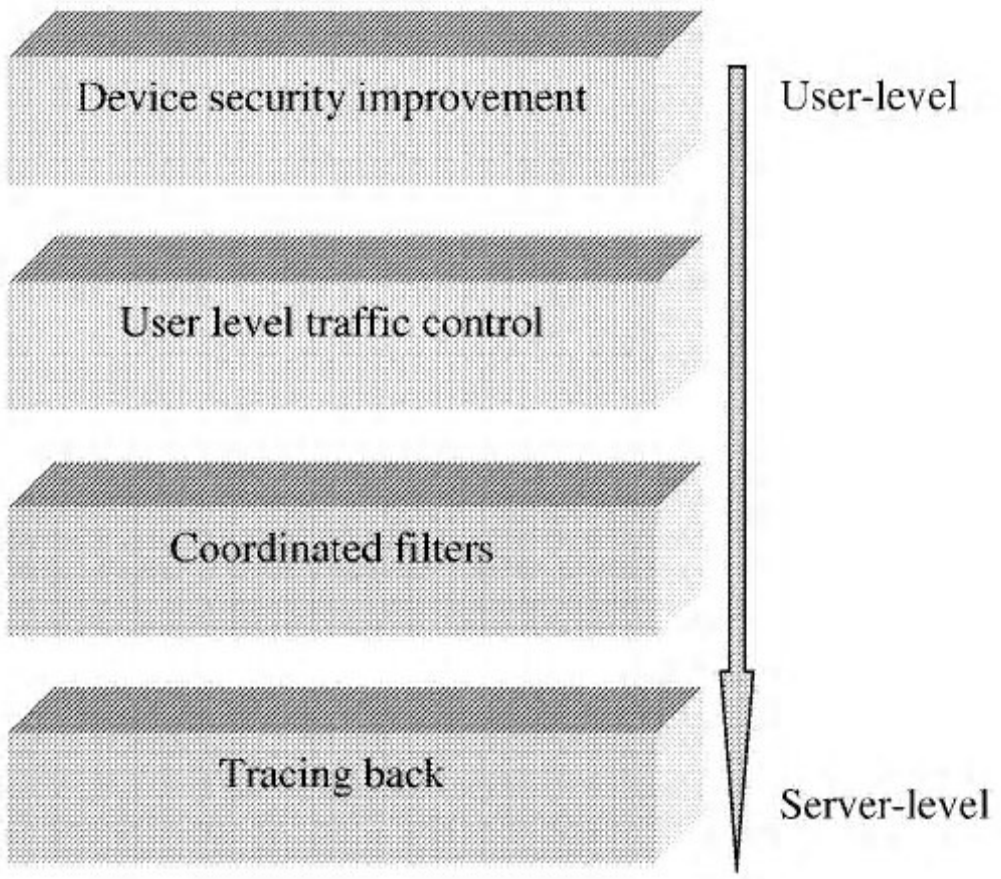

Figure 2.5. Four coordinated technological solutions to DDoS attacks.

Improving the security of all relevant devices

Before initiating an effective DDoS attack, the attacker needs to break into enough zombie devices to secure an ability to generate sufficient traffic. A direct counterstrike is to secure all devices to make it difficult for the attacker to seize enough zombies.

It is not practical, nor potentially beneficial, to secure all computers on the wired Internet. Alternatively, an effective and efficient solution would be to selectively secure those computers that have high traffic throughput – such as routers – or high performance

and high bandwidth workstations so that the marginal benefit for each dollar spent on security is optimized. Moreover, for some networks that have the ability to audit real-time traffic, security measures can even be delayed until a DDoS attack actually happens, thus making them more targeted and therefore more efficient.

For the wireless Internet, such a selective security implies that wireless devices with high bandwidth connections, e.g., 3G devices, are the ones that should be safeguarded. We note that the wireless communication industry has a tighter security tradition than the wired Internet community, partially because of the relatively large communication spectrum and device costs.

User-level traffic control

User-level traffic control is embodied in a set of traffic control rules specifically for a given network device. For example, a wireless device user can set up a daily traffic cap that is high enough not to disturb her/his normal usage, while abnormally large traffic will be stopped. Furthermore, the abnormal traffic may trigger a warning to the user or to a network administrator for follow-up diagnosis. Traffic control rules can be contingent on factors including other users' usage status. For example, a user can specify her/his data to be dropped or delayed if the network is experiencing congestion.

Geng and Whinston (2000) propose to use an e-stamp model to control traffic even if user devices are hacked. A direct implication is that user-level traffic control rules for a specific network device need to be protected more securely than the network device itself since we do not want the attacker to modify the traffic control rules once she/he gets control of a network device. For the wired Internet, Geng and Whinston propose to save the rules in edge routers because routers, given their concentrated and limited functionalities, are relatively easier to protect than other computers.

For the wireless Internet, the candidate host for traffic control rules can be flexible. Unlike desktop computers that are normally anonymous with concealed identity, wireless devices – especially wireless phones – have unique IDs or PINs that are transmitted along with the data which cannot be tampered with. These IDs or PINs can be used to identify wireless devices. Furthermore, unlike desktop computers in which software programs can control and modify virtually all information including the traffic control information, wireless devices normally have restricted access functions that enable secure traffic control even if the wireless device is hacked.

Edge routers in the Wireless Extended Internet and gateways in the Wireless Portal Network are the ideal hosts for coordinating user-level traffic control rules. For example, if a user wants her/his data packets to be dropped when the outbound network of the wireless ISP is congested, the edge router has the ability to realize this requirement. The designation of a host for traffic control rule coordination is complicated in a Wireless Ad Hoc Network since no one party is more likely to be in a central position than another.

Coordinated filters and tracing back

Even when user-level traffic control fails, wireless ISPs in the Wireless Extended Internet can still try to defeat DDoS attacks by identifying the attacking traffics and stopping them by using coordinated filters. The purpose of coordination among filters is to stop the traffic as early as possible along the attacking paths to prevent the damage from aggregated traffic. In a Wireless Portal Network, due to the relatively simple network topology, coordinated filters can be simplified to only one single filter. For a Wireless Ad Hoc Network, filtering is not applicable due to the symmetric structure. However, community rules, e.g., a voting mechanism, may play the role of a central filter to decide which user device to block.

Even if the coordinated filters cannot effectively stop the attack, possibly because the attacking traffic is hard to distinguish from normal traffic, there still exists another technological solution – to trace back to the zombie devices (and possibly the attacker) to shut down the attack from the source. Combining this with possible legal actions, this method can also help to deter repeated attacks.

A consistent incentive structure

According to the Yankee Group, a Boston consulting firm, the DDoS attack in February 2000 cost approximately $1.2 billion, not to mention the damage to consumer confidence in e-commerce (Murphy 2000). Effective coordinated solutions to DDoS attacks are critical for the future of e-commerce and m-commerce. However, a fervent advocacy of coordinated solutions does not necessarily result in actual implementation. Sample research by icsa.net, for example, shows that less than 15 percent of all corporate users are filtering source IP addresses. An even smaller percentage of Internet service providers – less than 8 percent – are doing this type of filtering (ICSA.Net 2000).

A disincentive structure for the wired internet

The reason for this low rate of implementation of coordinated solutions is the inconsistent incentive structure in Internet traffic pricing. Simply stated, the victim has the incentive to defend but cannot defend effectively, whereas the owners of zombie computers and ISPs can defend effectively but do not have the incentive to do so. In this time of flat monthly fee payments for wired Internet access, the owner of a zombie computer incurs little cost due to DDoS attacks since all that is stolen is just some traffic. On the other hand, preventing a personal computer from being controlled by any potential attacker requires frequent – virtually constant – monitoring and updating, at considerable cost. If the cost of protection is higher than the value of the traffic being protected, an

economic disincentive clearly exists. Similar logic applies to ISPs who can always collect the monthly fees no matter whether a DDoS attack happens or not. Thus, they may hesitate to install filters since they will lower network performance.

Who should be motivated to defend in the wireless Internet?

Having observed the failed incentive structure of the wired Internet, it is clear that the wireless infrastructure should contain a new incentive structure that can give wireless device owners and ISPs enough impetus to implement defense mechanisms. However, an efficient incentive structure need not target all wireless device owners – only high-bandwidth devices should be effectively protected, including:

- high-performance, high-bandwidth end-user devices (including wired devices that can communicatewith the wireless Internet),
- routers, and backbone switches.

As we mentioned before the possibility of attacking the Wireless Extended Internet from the wired counterpart, the incentive structure is also need for devices in the wired Internet. As wired devices generally have more communication capacity, the incentive structure for the wired network needs to be more strict.

According to MobileInfo.com (`http://www.mobileinfo.com/imode/buz_approach.htm`), the average monthly bill is $30–$40 (or ¥3700–¥4900). Therefore, part 2 is the leading cost.

Part 1: monthly charges	Part 2: packet transmission charges	Part 3: i-mode information charges
¥300/month	¥0.3 per packet (128 bytes)	Usually ¥100–¥300/month for each fee-based service

Table 2.1. i-mode pricing scheme (US $1 = Japan ¥123.5 as of July 12, 2001)

An incentive structure based on usage-based fees

One candidate for an effective incentive structure is the usage-based fee. The direct effect of a usage-based fee is a sharp increase in the cost to zombie devices if they are sending out attacking traffic. In particular, if a proper fee increase scheme is devised, it should not affect normal network usage but the cost could increase significantly for high-performance, high bandwidth devices when they are sending out huge traffic volume.

These computers are most often located in corporations, governments, and universities. With a usage-based fee structure, the owners of such computers will have the greatest immediate incentive to take security actions. Similarly in the wireless Internet, devices that have the potential to occupy a large portion of the radio frequency will be controlled most tightly. Likewise, a usage-based fee between an ISP and a backbone provider encourages the ISP to have more concern over its traffic. Specifically, such a usage-based fee plan makes ISPs more likely to install coordinated filters and to support user-level traffic controls.

Fortunately and unlike the wired Internet industry, the wireless Internet industry starts with usage-based fees. For example, Japanese vendor DoKoMo's i-mode service pricing is mainly packet based, as shown in Table 2.1.

US wireless providers are using minute-based pricing plans that are often simplified (as we will explain shortly) to the form of fixed pricing with an over-the-cap penalty for several service levels. Currently given the low bandwidth and simple functions of wireless devices in the US, simple pricing schemes based on connection time are applicable. However, it is conceivable that with the increase of bandwidth and more rich applications with different traffic requirements, and more importantly with the

migration to packet-based communication, packet-based pricing will become more accurate and practical than minute-based pricing.

The wireless Internet: Towards dynamic usage-based fees.

If the usage-based fee continues in the wireless Internet, we can expect less DDoS attacks compared with the wired Internet. A usage-based fee can be further calibrated to provide more targeted incentives against DDoS attacks, i.e., a dynamic usage-based fee plan can better prevent DDoS attacks than constant usage-based fees (CREC, Goldman Sachs 2000). A constant usage based fee scheme has a fixed unit price. Packet-based pricing is an example of the constant usage-based fee, while the dynamic usage-based fee implies a changing unit price, which is higher when there is congestion in the network (Goldman Sachs 2000).

Wireless service providers (as well as long-distance phone providers) have already considered predictable congestion for their constant usage-based fee scheme. For example, it is a common practice to price higher for daytime communication than for nighttime or weekend communication as congestion is more likely to happen in daytime. We call this the modified constant usage-based fee scheme.

A dynamic usage-based fee scheme, on the other hand, deals with unpredictable congestions, including those caused by DDoS attacks. The characteristic of a dynamic usage based fee is the increase in unit price when congestion happens or will happen. The incentive it gives to wireless device owners is twofold. First, those owners are more likely to set up traffic control rules in their device to instruct to delay or cancel the data transmission when the network is congested or approaching congestion. Therefore, even if an attacker instruct all zombie devices to send attacking traffic at the same time, an effectively synchronized attack is unlikely to occur. Second, as congestion means higher

cost, high bandwidth owners are more likely to invest more in the security of their devices to avoid stolen traffic.

Fee scheme	Characteristics	Effects	Examples
Constant usage-based fee	Fixed unit (unit traffic volume or unit communication time) price	Provides basic incentive for wireless device users to prevent traffic stolen	i-mode packet-based pricing
Modified constant usage-based fee	Multiple fixed unit prices, each for a given time period	Prevents/alleviates predictable congestions	Price differentiation for daytime and nighttime (e.g., SprintPCS)
Dynamic usage-based fee	Unit price increases when congestion happens or will happen	Prevents/alleviates any possible congestions	N/A

Table 2.2. Different usage-based fee schemes.

Fee scheme	Examples	Comments
Constant usage-based fee	Minute-based fee (many long-distance services in US), packet-based fee (i-mode)	The most preliminary usage-based fee scheme. Users are exposed to financial risks as they may receive large bills.
Constant usage-based fee with a cap	Pre-paid phone cards	An upper cap prevents financial risks for users. Nevertheless, the cap may be reached when important communication is going on.
Flat monthly fee with a communication cap	The next scheme is similar to this one. To our knowledge, now most plans allow over-the-cap usage for a higher price.	Limited communication volume/time implies that this is still usage-based pricing. Often several plans of different caps are offered to let users to self-select. Reduces the complexity of usage-base fee.
Flat monthly fee with over-the-cap penalty	Most cell phone plans from most US providers	High over-the-cap penalty (often around \$0.5/min) effectively stimulates usage control.

Table 2.3. Variations of constant usage-based fees.

Table 2.2 gives a concise comparison of three usage-based fee schemes.

Usage-based fees can be flexible

It is constantly questioned whether or not users will accept a usage-based fee plan even when it is financially beneficial for them. Some researches (Odlyzko2001) show that many people dislike the uncertainty and complexity associated with usage-based fees. Concerning this problem, it is worth pointing out that a consistent incentive structure can be flexible in its form while still representing the essence of a usage-based fee plan, as illustrated in Table 2.3.

For the Wireless Ad Hoc Network, a monetary incentive structure may not be available simply because of the lack of a charging system. Instead, other incentive mechanisms, e.g., a voting mechanism which effectively rules out a member upon heavy radio frequency usage, can serve the same purpose.

Once again, for defending the Wireless Extended Internet, a usage-based fee plan is also needed for the wired Internet. Nevertheless a usage-based fee plan for the wired Internet is mainly used to prevent DDoS attacks inside the wired Internet, for which Geng and Whinston (2000) have discussed possible mechanisms.

Cost-effectiveness

The history of the Internet shows that the de facto criteria for success in any proposal are whether that solution is proactive and consistent with mainstream and commercial Internet technologies. Because of the anonymous and "best effort" usage of the Internet, it is arduous and costly to regulate the infrastructure against DDoS attacks. Several advanced network management technologies have been proposed to address the traffic control problem. Employing these existing technologies will significantly reduce the costs and risks in designing future wireless Internet.

The Policy Based Networking (PBN) (Yavatkar et al. 2000) is one promising technology for implementing usage-based fees to deal with DDoS attacks. Essentially, it provides rules that describe actions to take when specific conditions arise. These policies are able to control critical network resources such as bandwidth, QoS, security and Web access across heterogeneous networks. Thus, both natural and artificial congestions are under the control of a globally coordinated structure. As illustrated in Figure 2.6, we present an implementation scheme based on the PBN, and discuss how to incorporate both the incentive structure and the technological solutions into this scheme in a cost-effective manner.

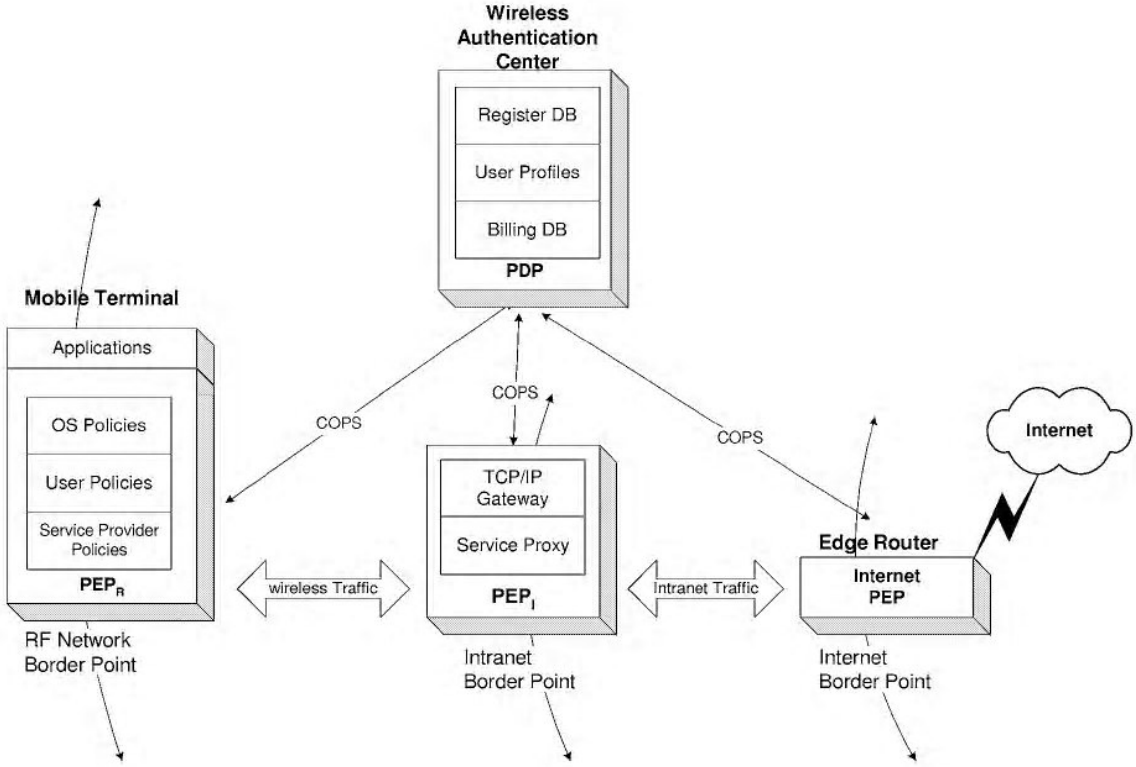

Figure 2.6. A wireless network architecture based on the PBN

In this scheme, the two main elements for policy control are the Policy Enforcement Point (PEP) and the Policy Decision Point (PDP) (Yavatkar et al. 2000). From the PBN perspective, the Wireless Location Register/Authentication Center is a natural policy server (i.e., PDP) with additional functionality such as user authentication, accounting, and policy information storage. At network border points, PEPs act as a "police" to accept or deny requests appropriately. Through secure and reliable channels (such as telecommunication out-of-band signaling network), PDPs and PEPs can exchange policy information with the Common Open Policy Service protocol (COPS) (Boyle et al. 2000).

At the user's end, with the Intelligent IC card and other hardware technologies, wireless devices have some embedded functionalities that cannot be tampered with. The

user-end policies have three levels. First, providers can deploy policies in terminals which users cannot change. Unlike desktop computers that are normally anonymous in the sense that they can conceal their identities, wireless devices such as wireless phones have unique IDs, or PINs, that are transmitted along with the data and cannot be altered. These IDs or PINs are effective instruments to identify wireless devices. Also, there are restricted access functions, such as integrating admission control into lower layer traffic control to increase the performance and security (Das et al 2000). These restrictions can enable secure traffic control of all relevant devices even if these devices are hacked.

Second, end users could design their own policies, which are unchangeable by applications. For example, a user can assign a daily cap in traffic for her/his cellular phone. If the cap is reached, the system could block any further transaction and/or raise an alarm. In fact, the pre-paid cellular phone card implements a similar traffic-cap function. Future mobile phone users can set rules that are more sophisticated.

The above two policy controls cannot be realized without specific hardware that is configurable only by providers or end users. A third level policy control can be constructed in software by enabling a wireless operating system to have multiple security levels. Policy control is realized in higher security levels that normal networking applications cannot modify.

Finally, at the Intranet border point, TCP/IP gateways play the role of policy proxies. Proper policy rules can turn these proxies into coordinated filters and even support advanced usage-based fee schemes, such as dynamic pricing. The entities involved in policy control can verify each other's identity and establish necessary trust links before communicating. With the help of standard PEPs on Internet edge routers, a global coordinated network will be formed to minimize theft and DDoS threats.

A usage-based fee scheme can be implemented by using PDPs and PEPs, for example, in the following way. First, once the fee scheme is decided, it is implemented as a set of policies in PDPs at the Wireless Authentication Centers. Secondly based on the fee scheme and the real-time traffic condition, a PDP decides the pricing rules for every related mobile terminal and send these rules as policies to PEPs on these mobile terminals. Thirdly PEPs on mobile terminals enforce these pricing rules. Whenever there is a surge in traffic, possibly caused by DDoS attacks, PEPs report the traffic change and any possible congestion to the coordinating PDP, who in return dynamically adjusts pricing rules according to the given fee scheme and instructs PEPs to update their pricing rules.

CONCLUDING REMARKS

The DDoS attack threatens all time-sensitive m-commerce services. Fortunately the wireless Internet currently has a distinctive advantage over the wired Internet in defending against the DDoS attack: the timing. When DDoS attacks came to the wired Internet, the infrastructure of the wired Internet had been stable for decades, albeit lacking reliable mechanisms for QoS control and incentive structures for traffic control. As a result, it was repeatedly targeted by DDoS attacks. In comparison, the wireless Internet industry has a chance to address DDoS attacks before it fully matures. However, time is running short as a well-founded wireless Internet infrastructure is expected to emerge by 2003 (Goldman Sachs 2000). Whether potential DDoS attacks on the wireless Internet will materialize or not will solely depend on how the wireless industry deals with the potential problem when solutions can still be embedded into the basic infrastructure.

CHAPTER 3: THE DYNAMICS OF ONLINE PEER-TO-PEER COMMUNITIES: AN EMPIRICAL INVESTIGATION OF MUSIC SHARING NETWORK AS AN EVOLUTIONARY GAME

Online peer-to-peer communities and online social network have become increasingly popular. In particular, the recent boost of online peer-to-peer communities leads to exponential growth in sharing of user-contributed content which have brought profound changes to business and economic practices. Understanding the dynamics and sustainability of such peer-to-peer communities has important implications for business managers. In this study, we explore the structure of online sharing communities from the dynamic process perspectives. We build an evolutionary game model to capture the dynamics of online peer-to-peer communities. Using online music sharing data collected from one of the IRC Channels for over five years, we empirically investigate the model which underlies the dynamics of the music sharing community. Our empirical results show strong support for the evolutionary process of the community. We find the two major parties in the community, namely sharers and downloaders, are influencing each other in their dynamics of evolvement in the community. The combination of the dynamics reveals the mechanism through which peer-to-peer communities sustain and thrive in a constant changing environment.

INTRODUCTION

The recent explosive growth of popular online peer-to-peer sharing communities (e.g., YouTube for video sharing, Flickr for photo sharing, and Digg for news sharing) has generated a renewed interest in the Internet as a new medium for content generation and distribution. This new trend is often considered to be attributable to the Web 2.0 technologies (e.g. Ajax, XML, RSS, and Wiki) and social computing concepts (e.g. blogging, tagging, and voting) that make mass user interactions feasible and multi-

faceted. Online sharing communities feature large populations of participants and constant change of community memberships. The communities also develop and sustain mainly by themselves without any corporate or commercial sponsorship. Individual users self-organize and construct communities on the Internet through large-scale collaboration and information sharing, which bring significant changes to e-commerce practices and organizational computing. Therefore, there is an increasing interest in understanding the structure and development of online social communities.

In particular, online sharing communities have become a venue where users directly provide services and products to each other, providing a new form of market for consumers. The development of online communities, however, challenges the traditional economic notion of market as participants merely are driven by costs and benefits with the sole objective of maximizing profits. In online peer-to-peer sharing communities, users and providers interact with thousands of fellow community members with limited knowledge of counterparts. Community participants often have few message exchanges and there are also few pre-existing social ties among them. Moreover, most member interactions among very large community populations are short-lived, and the community is dynamic with constant changes. These differences require a new approach for better understanding of the development and sustainability of online sharing communities.

To model these dynamic individual interactions with limited information in online communities, in this paper we adopt the evolutionary game approach in order to accurately portray the dynamics and evolutionary process of online peer-to-peer communities. Evolutionary game models (Samuelson 1997) emphasize large populations, continuous changes in community memberships, and imperfect information and memory among community members. These models are particularly powerful in interpreting users' behavior in online sharing communities (e.g. Geng et al. 2004). In

this research, we formulate a stylized evolutionary game model to capture the structural change and the development process of peer-to-peer communities. We then empirically test the evolutionary game model using over five years data collected from a major online music sharing community. As predicted by the evolutionary game model, our results suggest that the users in the community do follow an evolutionary path in terms of the dynamics of their music sharing activities. In particular, we find two dynamics which characterize the mechanism of development of sharers and downloaders accordingly. The combination of the two dynamics reveals the mechanism through which peer-to-peer communities sustain and thrive in a constant changing environment. We find that when sharer ratio decreases below certain threshold, downloaders exit and new sharers join the communities. This process gradually restores equilibrium to the peer-to-peer network. Similarly, when share ratio increases above certain threshold, sharers start to exit while new downloaders join the community. Remarkably, the dynamics also reveals the presence of a growth region where networks gain on both downloaders and sharers simultaneously. Our empirical analysis quantifies the speed at which peer-to-peer networks evolve and recover. We also demonstrate the similarity between the evolutionary game approach and the disequilibrium approach that has been used in prior studies. We show both approaches provide a structured understanding of the dynamics of peer-to-peer networks and product similar empirical results despite their differences in underlying assumptions.

The remainder of this paper is organized as follows. Section 2 summarizes related literature on online communities. Section 3 provides an overview of online peer-to-peer sharing community and discusses the advantages of evolutionary game in modeling social interaction and social communities. Section 4 constructs a stylized evolutionary game model that motivates our empirical analysis. Section 5 tests the model

empirically using data from IRC music sharing communities. Section 6 compares the similarities and differences between the evolutionary game approach and the disequilibrium approach. Section 7 concludes the paper with a discussion of the results and implications as well as identifying future research opportunities.

LITERATURE REVIEW

Our paper draws two main streams of literature that explore the dynamics of online communities. One stream investigates online communities at the aggregate level, while the other focuses on examining the communities at the individual level.

Extant research taking the aggregate approach view online communities as a whole and examine the impact of overall community activities on the growth of the community. Among the first systematic studies, Butler (2001) examined the role of communication activity on membership size using data collected from Internet ListServs. He proposed a resource-based model that treats membership size and communication activity as resources and benefits of the community. The model also recognizes that the large number of participants and communication activities in the community may also incur costs to members. He found that as membership size grows, the community experiences a faster "churn" rate, i.e., the percentage of membership loss increases. The results suggested that while more community activities provide more value to members in general, the net benefit does not increase monotonically. He thus cautions theorists and developers of online social structures to be aware of the opposing forces and the endogenous nature of membership size and communication activity, as well as their interplay, and adjust their expectations of the growth of a community accordingly.

Asvanund et al. (2004) are among the first to empirically study costs and benefits in peer-to-peer (P2P) music-sharing communities. As with Butler (2001), they recognize that as the network grows, so do the benefits (more resource availability) and costs

(network congestion due to user free riding). By sending queries to six P2P networks, they collected query congestion, song availability and download delay data to measure the exact value of positive and negative network externalities. Although in a different context, their result is similar to Butler's (2001) in that the marginal value brought by an additional user declines in larger networks, while the marginal costs imposed by the new user increases with size.

Extant literature taking the aggregate approach illustrates the potential impact of overall community activities on the development of community. However, the underlying mechanisms driving individual participation in the community remain unclear. Another stream of literature on sharing incentives focuses on analyzing individual rationales to participate and contribute to online communities. Jones et al. (2004) studied the effect of individual messages and information richness in online forums on user interactions (replies) and propensity to stay. The results suggest that simpler messages may encourage more active participation from users. Based on the classical economic theory, researchers assume that individuals maximize their direct payoff in deciding whether to stay in a community (Asvanund et al. 2004). However, in addition to the direct payoff through messages exchange, files downloaded, and traffic redistribution (Krishnan et al. 2004), it is suggested that individuals may join the community because of their own psychological and social characteristics, such as satisfying user needs (Raymond 2000, Lakhani and Wolf 2005), reciprocity (Kollock 1999, Shah 2006) and altruism (Torvalds 1998, Hars and Ou 2001).

In this paper, we combine the aggregate approach with the individual approach to model the dynamic evolution of a peer-to-peer network. Our approach complements these earlier studies by focusing on the sustainability and dynamic changes in these networks.

PEER-TO-PEER COMMUNITIES AND EVOLUTIONARY GAME APPROACH

The continual growth of the Internet and telecommunication networks boost up the recent development of peer-to-peer online sharing communities. As a relatively new phenomenon, online peer-to-peer sharing communities are characterized by their user-centered, content-based, quickly-expanded, and loosely-connected structure and development. The growth of online communities challenges the traditional economic notion of the individual as a payoff maximization agent interacting with the market with full information and complete rationality. In online communities, individuals no longer deal with products or services provided by a few monopolistic firms. Instead, products (such as music and photo files) are provided by other members voluntarily in the community. Users in the community are both providers and consumers. As such, individual's behavior and community development is influenced by the overall provision and consumption activities and the interactions among community members in the community. In this research, we are particularly interested in exploring what are the underlying mechanisms of the dynamics and development of such user-centered communities.

From social network and communication perspectives, we have limited understanding of the structure and dynamics of online social communities because of the highly dynamic and ad-hoc natures of interactions among members. On the Internet, people interact with each other for information and content and they often have few pre-existing connections. In some cases, such as music/video sharing and Open Source Software (OSS) collaboration, they may not even have any direct communication. Instead of mutual social connections that link individuals together (Monge and Contractor 2003, pp. 156), the most important impact an online peer-to-peer community has on its members is through the individual's observation of system characteristics and

aggregate behavior of all other users. For example, in digg.com, popular ranking composed by readers' votes determines the importance level for each piece of news and influences others' reading preferences. In order to understand these dynamics, we need to go beyond the traditional social network approach that focuses on characterizing the interactions and relationships among community members.

Based on classical game theory, current research often assumes that individuals are fully rational about others and the market. These assumptions cannot be applied to the user-centered online sharing communities. In online peer-to-peer communities, heterogeneous members may have different and limited rationality. Nonnecke et al. (2006) suggest that individual behavior (in MSN communities) is temporary and usually adapts to exogenous and endogenous factors. In this scenario, the traditional game theory assumptions have limited power to characterize the dynamics of individual strategies.

Unlike traditional game theory models, which assume that all players are fully rational and have complete knowledge of others and the market, three advantages make evolutionary games a preferable approach to study collective and dynamic influence in online user-centered communities. Firstly, evolutionary game theory assumes that people decide their strategies gradually and adaptively. They do not have to be rational and optimal. Through trial and learning processes, low-payoff strategies tend to be weeded out and equilibria may emerge (Samuelson 1997). This selection mechanism captures the formation of collective behavior in online communities. Secondly, evolutionary game theory has the ability to model changing population – one of the most important merits from the marriage of economics and biology. This makes the evolutionary approach powerful in studying large and open communities. Without restrictions on the number of players, evolutionary models can describe many membership dynamics, such as member

gain, loss, and dominance. Finally, evolutionary game models focus on the convergence process from far-from-equilibrium instead of steady states near-equilibrium.

In this paper, we take the initial step to investigate the underlying mechanism of the dynamics of online sharing communities using data collected from a major music sharing network. Music sharing communities provide an excellent context for this study as music sharing network are by far the most popular online communities that have attracted the largest number of users (Asvanund et al. 2004). Next section, we formulate a basic evolutionary game to characterize the dynamics of an online music sharing community with two major types of users, that is, the sharers who provide music files and the downloaders who search, request, and download music files.

ONLINE SHARING COMMUNITIES AS EVOLUTIONARY GAMES

Evolutionary game theory views a community as a collection of interactions between individuals over time. The payoff for an individual in a given time period is determined by her strategy and the strategy of her counterpart, as in the traditional game theory framework. Evolutionary game theory however differs from the traditional game theory with regard to its assumption of individual rationality. Instead of having individuals choose strategies based on perfectly rational expectation of the future, evolutionary game theory imposes a lower requirement. It assumes individuals are myopic and their choices of strategy are influenced only by their payoffs in the most recent time period (Taylor and Jonker 1978). This process is called replicator dynamics. The lower requirement on rationality also suggests that individuals do not reach optimal strategy instantly. Rather, it is a gradual process played out over time.

Population and payoffs in sharing communities

In online sharing communities, we assume an individual makes two decisions. First, she decides whether she would like to share music files with other users of the peer-to-peer network. Second, she decides whether she would like to download music files from the network. Individuals who decide to share music files are called sharers, while those who decide to download are called downloaders. Evolutionary game theory requires the two decisions to be considered separately. That is, sharers only provide resources to others while downloaders only download without contribution. If an individual both share and download music files, the two decisions are assumed to be independent from each other. Given the independence assumption, an individual that shares and downloads music files at the same time can be modeled as two individuals, one as a sharer and the other as a downloader. We denote sharer's subpopulation at time t as $x_S(t)$ and downloader's subpopulation as $x_F(t)$.

Both sharers and downloaders derive payoffs from the online peer-to-peer network. Sharers take joy in sharing his or her music collection with others. Such joy could come from community status, influence and self-perception. The payoff received by each sharer may vary. Sharers whose music files have been downloaded more frequently may receive a higher payoff than those whose files are downloaded less frequently. Similarly, downloaders receive payoff in obtaining music downloads. Those who have obtained more downloads could receive higher payoff than others. To model the behavior of the population that consists of such heterogeneous individuals, evolutionary game theory assumes that the population is large and the interactions between individuals are random[1]. Given the assumption, the behavior of the population

[1] This assumption fits well with peer-to-peer music sharing networks where downloaders and sharers are anonymous and have no social interaction except for music downloading.

is contingent upon average payoffs of the two types of individuals. We denote sharers' average payoff at time t as $v_S(t)$ and downloaders' average payoff as $v_F(t)$. Sharer's average payoff is influenced by number of downloads requested from each sharer, which is in turn determined by the proportion of downloaders in the community. $v_S(t)$ can therefore be considered as a function of downloader ratio $x_F/(x_S + x_F)$. Likewise, downloader's average payoff is influenced by number of downloads they obtained from sharers, which is in turn determined by the proportion of sharers in the community. Therefore, we can express $v_F(t)$ as a function of sharer ratio $x_S/(x_S + x_F)$. The above discussion also indicates that a peer-to-peer network is uniquely determined by its population state $(x_S(t), x_F(t))$. Given the population state, sharers and downloaders realize their payoffs $v_S(t)$ and $v_F(t)$, which determine the change of subpopulations in the next period. This process creates a path of evolution.

Population Dynamics

Sharer and downloader subpopulations changes through gaining new users and losing existing users (also called birth rate and death rate in the evolutionary game theory). Based on *replicator dynamics*, user gain and user loss are determined by their payoffs. When payoff increases, the community gains more new users and loses less existing users. For tractability, evolutionary game theory typically assumes the rate of change is a linear function of the payoff. For example, when $x_S(t)$ sharers have $v_S(t)$ payoff in period t, evolutionary game theory assumes the net change rate $\frac{dx_S(t)}{x_S(t)dt}$ equals to $a_S v_S(t) + b_S$, where a_S represents the degree to which sharer population is influenced by the average payoff it receives and b_S captures the inherent growth of in number of sharers. We can further decompose the change in sharer population $dx_S(t)$ into gain of

---

It however may not fit with other types of social networks where individuals may develop social ties that lead to repeated interaction with each other.

new sharers $dx_S^g(t)$ and loss of existing sharers $dx_S^l(t)$ where $dx_S(t) = dx_S^g(t) - dx_S^l(t)$. The decomposition provides more details of the dynamics how the sharer population in a peer-to-peer network evolves. We therefore propose

$$\frac{dx_S(t)}{x_S(t)dt} = a_S v_S(t) + b_S \tag{1}$$

$$\frac{dx_S^g(t)}{x_S(t)dt} = a_S^g v_S(t) + b_S^g \tag{1a}$$

$$\frac{dx_S^l(t)}{x_S(t)dt} = a_S^l v_S(t) + b_S^l \tag{1b}$$

Equations (1a) and (1b) indicate sharer subpopulation reaches its stable point when gain in new sharers $a_S^g v_S(t) + b_S^g$ equals the loss of existing sharers $a_S^l v_S(t) + b_S^l$. Let v_S^* be sharer payoff at the stable point. When the network offers higher payoff to sharers, the peer-to-peer network obtains a net gain of sharers. Otherwise, the network incurs a net loss of sharers.

We apply the same analysis to downloaders. We model the net change rate of downloaders $\frac{dx_F(t)}{x_F(t)dt}$ as a linear function of average downloader payoff $a_F v_F(t) + b_F$, where a_F represents the degree to which downloader subpopulation is influenced by the average payoff it receives and b_F captures the inherent growth of number of downloaders. Again, we have

$$\frac{dx_F(t)}{x_F(t)dt} = a_F v_F(t) + b_F \tag{2}$$

$$\frac{dx_F^g(t)}{x_F(t)dt} = a_F^g v_F(t) + b_F^g \tag{2a}$$

$$\frac{dx_F^l(t)}{x_F(t)dt} = a_F^l v_F(t) + b_F^l \tag{2b}$$

Differential equations in Equations (1) and (2) capture the complete dynamics for the evolution of the peer-to-peer networks. The objective of this study is to empirically

validate the dynamics proposed by evolutionary game theory and to quantify the evolution process of peer-to-peer networks. The empirical validation of the dynamics requires us to revise Equations (1) and (2) to discrete time periods. We therefore have

$$\begin{cases} \dfrac{x_S(t)-x_S(t-1)}{x_S(t-1)} = \left(a_S v_S(t-1)+b_S\right) \\ \dfrac{x_F(t)-x_F(t-1)}{x_F(t-1)} = \left(a_F v_F(t-1)+b_F\right) \end{cases} \tag{3}$$

$$\begin{cases} \dfrac{x_S^g(t)-x_S^g(t-1)}{x_S(t-1)} = \left(a_S^g v_S(t-1)+b_S^g\right) \\ \dfrac{x_S^l(t)-x_S^l(t-1)}{x_S(t-1)} = \left(a_S^l v_S(t-1)+b_S^l\right) \\ \dfrac{x_F^g(t)-x_F^g(t-1)}{x_F(t-1)} = \left(a_F^g v_F(t-1)+b_F^g\right) \\ \dfrac{x_F^l(t)-x_F^l(t-1)}{x_F(t-1)} = \left(a_F^l v_F(t-1)+b_F^l\right) \end{cases} \tag{4}$$

Equation (3) describes the dynamics with regard to net changes in the two sub-populations, while Equation (4) provides detailed dynamics on gain and loss of sharers and downloaders.

EMPIRICAL ANALYSIS

Evolutionary game theory provides an elegant explanation of the process by which economic agents in a community converge to equilibrium behavior without requiring such agents to have full information and perfect foresights about the market. A direct implication of the theory is that agents converge to equilibrium over time and the speed of the convergence is determined by the mechanism used by the agents. Based on the theoretical development discussed in the previous section, we empirically test the evolution of online sharing community using data collected from a music sharing community in the IRC Undernet (Mp3passion).

Data Description

We use data from the Mp3passion channel – a music sharing community in the IRC Undernet – to test the implication of evolutionary game. From March 2001 to May 2006, we monitored and recorded all user activities in the community. On average more than 56,000 files were downloaded per day in this community, equivalent to 0.05% of the global music sharing volume (Wall Street Journal, 19 November 2003). Our data provide detailed information on sharers and downloaders at different time periods. In addition, our data also capture changes in user types, sharing and downloading activities. The wealth of the dataset enables us to examine the evolution and dynamics of the community.

Our log data of user activities are aggregated on the daily basis. Tables 3.1 and 3.2 show the descriptions and the descriptive statistics for the key variables used in this paper. The summary statistics suggest that on average 3539 downloaders using the peer-to-peer network daily and 2399 among them are new comers in a given day. Despite the significant number of newcomers, the downloader population is overall stable as about the same number of downloaders exit from the network daily. We observe similar phenomenon for sharers with an average of 526 sharers use the peer-to-peer network daily, 203 of which are newcomers in a given day. The overall sharer population is also stable with equal number of sharers joining and exiting the network daily. Table 3.2 also shows that sharers account for about 13% of the total population with a standard deviation of 4%.

Variable	Description
Downloader(t)	The number of users who only download music at day t
DownloaderGain(t)	The number of new downloaders observed at day t
DownloaderLoss(t)	The number of downloaders disappeared at day t

Sharer(t)	The number of users who share music at day t
SharerGain(t)	The number of new sharers observed at day t
Sharer Loss(t)	The number of sharers disappeared at day t
SharerRatio(t)	Proportion of sharers in the population at day *t*

Table 3.1. Variable Description

Variable	*N*	*Mean*	*Median*	*SD*	*Min*	*Max*
Downloader(t)	1001	3539.64	3580	355.92	1386	5164
DownloaderGain(t)	1001	2399.93	2410	266.90	835	3841
DownloaderLoss(t)	1001	2402.47	2422	261.27	959	3866
Sharer(t)	1001	526.30	548	174.10	154	1440
SharerGain(t)	1001	203.79	197	110.94	41	1025
Sharer Loss(t)	1001	203.87	195	112.44	35	1047
SharerRatio(t)	1001	0.13	0.14	0.04	0.04	0.31

Table 3.2. Summary Statistics of Daily Data

Based on the download requests and associated user IDs in the raw data, we can build a downloader list and count the number of unique individuals who only download files in day t as *Downloader(t)*. Downloader gain (loss) was calculated by comparing each day's list to the previous day's to determine the number of new downloaders (or downloaders disappeared). Similarly because each file server should announce its status regularly, we can compose a list of sharers and calculate *Sharer(t)*, *SharerGain(t)*, and *SharerLoss(t)*.

Empirical Model

Using the IRC data, we calculate daily rates of net gain (loss), absolute gain and absolute loss of sharers and downloaders. These daily change rates are dependent variables in equation (3) and (4). The independent variable in equation (3) and (4) is average payoff to sharers and average payoff to downloaders. As we mentioned earlier, evolutionary game theory assumes a large population and individuals in the network have

equal opportunity to meet each other in each time period (Ellison 1993). Given the assumption, we can write the average sharer payoff as a linear function of the proportion of downloaders in the network $x_F/(x_S+x_F)$. Similarly, the average download payoff can be expressed as a linear function of the proportion of sharers in the network $x_S/(x_S+x_F)$. Note that $x_F/(x_S+x_F)=1-x_S/(x_S+x_F)$, the payoffs to sharers and downloaders can be fully summarized by the sharers' ratio. Given that payoff functions are equivalent up to positive linear transformation, we have

$$v_S(t)=\frac{-x_S(t)}{(x_S(t)+x_F(t))} \text{ ; and} \tag{5}$$

$$v_F(t)=\frac{x_S(t)}{(x_S(t)+x_F(t))} \tag{6}$$

Substituting equation (5) and (6) into Equations (3) and (4) and adding the necessary noise term, our estimation model is as follows:

$$\begin{cases} \dfrac{x_S(t)-x_S(t-1)}{x_S(t-1)}=\left(\dfrac{-a_S x_S(t-1)}{(x_S(t-1)+x_F(t-1))}+b_S\right)+\varepsilon_t \\ \dfrac{x_F(t)-x_F(t-1)}{x_F(t-1)}=\left(\dfrac{a_F x_S(t-1)}{(x_S(t-1)+x_F(t-1))}+b_F\right)+\varepsilon_t \end{cases} \tag{7}$$

$$\begin{cases} \dfrac{x_S^g(t)-x_S^g(t-1)}{x_S(t-1)}=\left(\dfrac{-a_S^g x_S(t-1)}{(x_S(t-1)+x_F(t-1))}+b_S^g\right)+\varepsilon_t \\ \dfrac{x_S^l(t)-x_S^l(t-1)}{x_S(t-1)}=\left(\dfrac{-a_S^l x_S(t-1)}{(x_S(t-1)+x_F(t-1))}+b_S^l\right)+\varepsilon_t \\ \dfrac{x_F^g(t)-x_F^g(t-1)}{x_F(t-1)}=\left(\dfrac{a_F^g x_S(t-1)}{(x_S(t-1)+x_F(t-1))}+b_F^g\right)+\varepsilon_t \\ \dfrac{x_F^l(t)-x_F^l(t-1)}{x_F(t-1)}=\left(\dfrac{a_F^l x_S(t-1)}{(x_S(t-1)+x_F(t-1))}+b_F^l\right)+\varepsilon_t \end{cases} \tag{8}$$

Results

Table 3.3 presents the estimation results. Column (1) and (2) shows the dynamics of net changes in sharers and downloader population. The corresponding equation forms are as follows.

$$DownloaderChangeRate(t) = 0.17(SharerRatio(t-1) - 0.11) \quad (9)$$

$$SharerChangeRate(t) = -0.21(SharerRatio(t-1) - 0.15) \quad (10)$$

where DownloaderChangeRate $= \frac{x_F(t) - x_F(t-1)}{x_F(t-1)}$, SharerChangeRate= $\frac{x_S(t) - x_S(t-1)}{x_S(t-1)}$ and SharerRatio= $\frac{x_S(t-1)}{(x_S(t-1) + x_F(t-1))}$.

Equation (9) suggests that downloaders are resource seeking and the change of downloader subpopulation is self-regulated: when the sharer ratio is low and downloaders get less payoffs in the community, some of them start leaving and the increasing sharer ratio improves the payoff for the rest of downloaders. Likewise, when the sharer ratio is high and downloaders get more payoffs, new downloaders join the community, leading to a decrease in sharer ratio. The equation shows that downloader subpopulation reach a rest point when sharer ratio equals to 11%, in which case there will be no change in number of downloaders.

On the other hand, as shown in equation (9), the change of sharer population is negatively correlated with the sharer ratio. This result is consistent with sharer altruism proposed by Torvalds (1998). Sharers are interested in contributing resources and their payoff increases with number of downloaders in the community. More sharers will join if the sharer ratio is lower than 15%. On the other hand, when there are more than 15% sharers in the peer-to-peer network, sharer subpopulation will decrease. That is, sharers' rest point is 15%.

Variable	*Equation (7)*		*Equation (8)*			
	DownloaderChange Rate	*SharerChange Rate*	*DownloaderGain Rate*	*DownloaderLoss Rate*	*SharerGain Rate*	*SharerLoss Rate*
SharerRatio(t-1)	0.16*** (0.06)	-0.21*** (0.07)	0.11* (0.06)	-0.061*** (0.02)	0.92*** (0.07)	1.13*** (0.04)
Constant	-0.019** (.004)	0.031*** (0.01)	0.67*** (0.008)	0.69*** (0.002)	0.26*** (0.01)	0.23*** (0.005)
*Note: *** p<.01 ** p<.05 * p<.10*						

Table 3.3. Direct Estimation Results

However neither 11% nor 15% is an equilibrium point for the peer-to-peer network. This is because downloaders and sharers interact with each other. Solving Equations (9) and (10) together, we find the network reaches an equilibrium sharer ratio at 13% at which both downloader and sharer subpopulations grow steadily at the same speed. Our results also show the existence of a growth region between 11% and 15% where the downloader population and the sharer population increase simultaneously. Figure 3.1 provides a graphic representation of the underlying dynamics of the peer-to-peer networks. The solid straight line requires the equilibrium sharer ratio at 13%. The two dotted straight lines on the two sides represent the growth region bounded by sharer ratio at 11% and 15% respectively. For network states lie within the growth region (Region III), the paths of population changes are featured by simultaneous increase in number of sharers and downloaders. However, for network states lie outside of the growth region (Regions I and II), their paths of population changes are characterized by first a decrease in one of the subpopulations and then simultaneous increases in number of sharers and downloaders. In either case, the population is stable because any sharer ratio will convert to the equilibrium ratio eventually. Our results explain the mechanism by which a peer-to-peer network grows and we show that the growth of the network is embedded in the complementary between downloaders and sharers. Our results also reveal the mechanism by which peer-to-peer networks avoid unstable dynamics such as

avalanche and collapse. We find that downloaders and sharers dynamics can self-recover when the networks deviates from equilibrium.

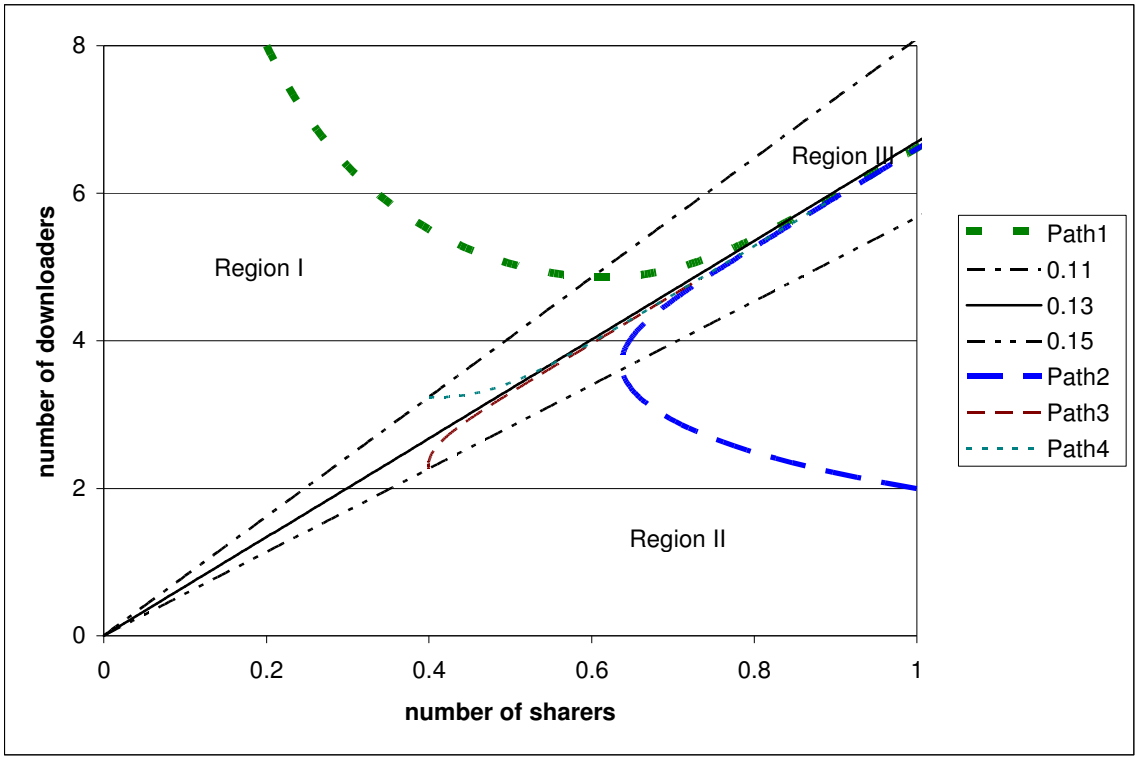

Figure 3.1. Sharer Ratios and the Paths of Population Changes

We further consider the process through which the peer-to-peer network gains or losses its users. Columns (3) – (6) in Table 3.3 estimate the dynamics for gain and loss of downloaders as well as gain and loss of sharers. The equations for the dynamics are as follows

$$DownloaderGainRate(t) = 0.11SharerRatio(t-1)) + 0.66 \quad (11)$$
$$DownloaderLossRate(t) = -0.06SharerRatio(t-1)) + 0.69 \quad (12)$$
$$SharerGainRate(t) = 0.92SharerRatio(t-1)) + 0.26 \quad (13)$$
$$SharerLossRate(t) = 1.13SharerRatio(t-1)) + 0.23 \quad (14)$$

The results show that downloaders have an inherent gain rate of 66% and loss rate of 69%, while the inherent gain and loss rate of sharers are 26% and 23% respectively. This suggests the overall turnout rate for downloaders is much higher than for sharers. We also find that the overall impact of sharer ratio on change in downloader population is relatively small. Each 10% increase in sharer ratio in the previous time period increases downloader gain by about 1% and decrease downloader loss by 0.6%. On the other hand, sharer ratio has a much bigger impact on change in sharer population.

EVOLUTIONARY GAME AND DISEQUILIBRIUM MODELS

Evolutionary game theory suggests the convergence to equilibrium is a gradual process due to myopic users. An alternative explanation of the dynamics is cost for changes. For example, a downloader who has exited a peer-to-peer network is unlikely to be aware of recent increases in number of sharers on the network. Such information disadvantage is a type of adjustment costs and reduces the speed of convergence to the equilibrium. Griliches (1967) and Maddala (1983) show that in a market with adjustment costs, agents may not fully adjust their strategies and the market equilibrium may reflect only a partial adjustment. In this section, we apply disequilibrium models to the study of evolutionary in peer-to-peer networks and compare the similarity between evolutionary game model and the disequilibrium model.

We again assume that the network state is determined by number of downloaders and number of sharers. Downloaders can be considered as the daily demand in the peer-to-peer network, while number of sharers can be considered as the daily supply in the network. The equilibrium between the demand and supply can be specified as follows:

$$\begin{cases} x_S(t) = a + bx_F(t) + \varepsilon_t \\ x_F(t) = c + dx_S(t) + \varepsilon_t \end{cases} \tag{15}$$

The interaction between the two equations determines the market equilibrium as defined in the traditional game theory literature. The assumption of disequilibrium model, however, indicates the equilibrium does not arise instantly. Due to adjustment costs, market equilibriums only reflect partial adjustments. To capture the partial adjustment process, the disequilibrium approach suggests Equations (15) shall be revised as follows:

$$\begin{cases} x_S(t) = x_S(t-1) + \lambda_1\left(a + bx_F(t) - x_S(t-1)\right) + \varepsilon_t \\ x_F(t) = x_F(t-1) + \lambda_2\left(c + dx_S(t) - x_F(t-1)\right) + \varepsilon_t \end{cases} \quad (16)$$

In the above model, λ represents the speed at which the market converges to the equilibrium. When λ equals to 1, equations (16) become the same as equations (15), indicating that users in the network can instantly reach the equilibrium and rejecting the disequilibrium model. The smaller λ goes, the longer it takes for the market to reach the market equilibrium, providing supports for the presence of adjustment costs.

We simultaneously estimate equations (16). The results (Table 3.4) show strong evidence that convergence to market equilibrium is not instant. The estimation for the sharer equation indicates a coefficient of 0.93 on $x_S(t-1)$, which corresponds to a λ_1 value of 0.07. The small value of λ_1 indicates a slow speed at which the market converges to the equilibrium. In each period, the market makes a partial adjustment of only 7% for sharer dynamics. The estimation for the downloader equation is slightly better. The coefficient on $x_S(t-1)$ is 0.70, representing a λ_2 value of 0.30. The results again suggest the market does not converge to equilibrium instantly. Rather, the market makes a partial adjustment of about 30% in each time period for the downloader dynamics. Overall, we find users in the peer-to-peer music sharing community take a long time to converge to market equilibrium and the convergence is mainly driven by the mobility of users who request files. They are more responsive to changes in availability of resource on the market.

Variable	Parameter Estimate	Variable	Parameter Estimate
Equation (16): Downloader equation		*Equation (16): Sharer equation*	
Downloader(t-1)	0.70*** (0.02)	***Sharer(t-1)***	0.93*** (0.01)
Sharer(t)	-0.16*** (0.05)	***Downloader(t)***	-0.03*** (0.008)
*Note: *** p<.01 ** p<.05 * p<.10*			

Table 3.4. Simultaneous Equation Estimation Results

CONCLUDING REMARKS

The blossom of online communities has intrigued both academic researchers and popular press. Understanding the dynamics and the sustainability of such communities has important implications in investigating their influences on e-commerce as well as contributing to the current online communities and online social network literature. In this paper, we take an evolutionary game approach to model online sharing communities from a dynamic perspective. More importantly, we take the first step to empirically test the evolution of the online sharing community using data collected from the IRC music sharing network. Our model predicts that the two major members (downloaders and sharers) in the community both play an essential role in the community. In contrast to the traditional view that downloaders only consume resources in the community, our model predicts that the existence of downloaders and their self-regulation effect influence the growth of sharers' population. Thus, the tendency of avalanche and collapse is reduced. In other words, downloaders act as a resistance or stabilizer of a community. Our empirical results show strong support for the evolutionary process of the community. In addition, our results indicate that users in the music sharing community take a long time to converge to market equilibrium and the convergence is mainly driven by the mobility of users who request files (mostly downloaders).

The future extensions of this research lie in the following directions. First, the evolutionary game model will be extended to incorporate other dynamics in order to capture more complex structure of online communities. Second, the empirical test will be fully extended to examine the behavior, activities, conversion, and transformation of various members in the community from a dynamic perspective. Third, the empirical test will also be extended to different types of online communities in order to explore the interpretive power of evolutionary game approach in online social networks.

Bibliography/References

Asvanund, A., Clay, K., Krishnan, R., and Smith, M. 2004. An empirical analysis of network externalities in peer-to-peer music sharing networks, Information Systems Research 15(2), pp. 155-174.

Ba, S., Stallaert, J., and Whinston, A.B. 2001. Research Commentary: Introducing a Third Dimension in Information Systems Design -- The Case for Incentive Alignment. Information Systems Research 12, 225-239.

Badishi, G., Keidar, I., and Sasson, A. 2004. Exposing and Eliminating Vulnerabilities to Denial of Service Attacks in Secure Gossip-Based Multicast. In Proceedings of International Conference on Dependable Systems and Networks (DSN'04), Palazzo dei Congressi, Florence, Italy, June 2004, 223-232.

Boyle, J., Cohen, R., Durham, D., Herzog, S., Raja, R., and Sastry, A. 2000. The COPS (Common Open Policy Service) Protocol, IETF RFC 2748, Proposed standard (January 2000).

Butler, B. S. 2001. Membership Size, communication activity, and sustainability: a resource-based model of online social structures, Information Systems Research 12(4), pp. 347-362.

Cavusoglu, H., Mishra, B. K., and Ragunathan, S. 2002. The Effect of Internet Security Breach Announcements on Market Value of Breached Firms and Internet Security Developers. Workshop on Information Systems and Economics Program, Barcelona, Spain, December 2002.

Chang, R.K.C. 2002. Defending against Flooding-Based Distributed Denial-of-Service Attacks: A Tutorial. IEEE Communications Magazine 40, 42-51.

Constant, D., Kiesler, S., and Sproull, L. 1996. The kindness of strangers: On the usefulness of weak ties for technical advice, Organization Science (7), pp. 119-135.

CREC, 1997. Congestion pricing more profitable than AOL pricing, Research report, http://crec.mccombs.utexas.edu/news/nr.html

CTIA, 2000. Background on CTIA's semi-annual wireless Industry survey, http://www.wow-com.com/pdf/wireless_survey_2000.pdf

Currier, K.M. 2000. Comparative statics analysis in Economics, World Scientific Publishing Co.

Das, S.K., Jayaram, R., Kakani, N.K., and Sen, S.K. 2000. A call admission and control scheme for quality-of-service (QoS) provisioning in next generation wireless networks, Wireless Networks 6 17–30.

Dennis, S. 2000. Mobile phones emerge as new virus target Kaspersky, Newsbytes.com, http://www.newsbytes.com/news/00/153195.html

Ellison, G. 1993. Learning, Local Interaction, and Coordination, Econometrica, (61:5), pp.1047-1071.

Ettredge, M., and Richardson, V. 2002. Assessing the Risk in E-commerce. In Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS'02) vol. 7, Big Island, Hawaii, January 2002, IEEE Computer Society Press, Los Alamitos, CA, 194.

Fasbender, A., Kesdogan, D., and Kubitz, O. 1996. Analysis of security and privacy in Mobile IP, in: 4th International Conference on Telecommunication Systems, Modeling and Analysis.

Fisher, D. and Callaghan, D. 2001. Microsoft attack raises concern over new DDoS variant, Yahoo! News (January 26, 2001) http://dailynews.yahoo.com/h/zd/20010126/tc/microsoft_attack_raises_concern_over_new_ddos_variant_1.html

Geng, X., and Whinston, A.B. 2000. Defeating Distributed Denial of Service Attacks. IEEE IT Professional 2, 36-41.

Geng, X., Gopal, R., Ramesh, R., and Whinston, A.B. 2003. Scaling Web Services with Capacity Provision Networks. IEEE Computer 36, 64-72.

Geng, X., Gopal, R., Ramesh, R., and Whinston, A.B. 2005. Capacity Provision Networks: Foundations of Markets for Internet Caching. Working paper.

Geng, X., Huang, Y., and Whinston, A.B. 2002. Defending Wireless Infrastructure Against the Challenge of DDoS Attacks. ACM Journal on Mobile Networking and Applications 7, 213-223.

Geng, X., Whinston, A. B., and Zhang, H. 2004. Health of Electronic Communities: An Evolutionary Game Approach, Journal of Management Information Systems (21:3), pp. 83-110.

Goldman Sachs, 2000. Technology: Mobile Internet, Research report (September 2000).

Gupta, A., Stahl, D.O., and Whinston, A.B. 1999. The Economics of Network Management. Communications of the ACM 42, 57-63.

Haas, Z.J., and Lin, Y. 2000. Demand re-registration for PCS database restoration, Mobile Networks and Applications 5, 191–198.

Haney, C. 2001. Network Associates hit by denial-of-service attack, IDG News Service (February 2, 2001) http://www.computerworld.com/cwi/stories/0,1199,NAV47-68-84-88-93_STO57290,00.html

Harvey, N. J.A., Jones, M.B., Saroiu, S., Theimer, M., and Wolman, A. 2003. SkipNet: A Scalable Overlay Network with Practical Locality Properties. In Proceedings of the Fourth USENIX Symposium on Internet Technologies and Systems, Seattle, WA, March 2003.

Hopper, D.I. 2000. Denial of service hackers take on new targets, CNN.com (February 9, 2000) http://www.cnn.com/2000/TECH/computing/02/09/denial.of.service/

Huang, Y., Geng, X., and Whinston, A.B. 2003. Network Mapping Services for Provisioning of Decentralized Web Services: Promises and Issues. In Proceedings of the 2nd Workshop on e-Business, Seattle, WA, December 2003.

ICSA.Net, 2000. 650-member Alliance for Internet Security unveils tool to detect network vulnerability, http://www.icsa.net/html/press_related/2000/3_23_00_NetLitmus.shtml

Internet Security Systems, Denial of Service FAQ, http://www.iss.net/news/denialfaq.php

Jones, Q., Ravid, G., and Rafaeli, S. 2004. Information overload and the message dynamics of online interaction spaces: a theoretical model and empirical exploration, Information Systems Research 15(2), pp. 194–210.

Ledyard, J.O., and Szakaly-moore, K. 1994. Designing Organizations for Trading Pollution Rights, Journal of Economic Behavior and Organization 25, 167-196.

Kaspersky Labs Int. demystifies the discovery of the first “true” wireless virus, http://www.avp.ru/news.asp?tnews=0&nview=1&id=107&page=0

Kleinbard, D. 2000. More sites hacked in wake of Yahoo!. CNN Money News (Feb. 8, 2000), Published on the Web, <http://money.cnn.com/2000/02/08/technology/yahoo>.

Kollock, P. 1999. The economies of online cooperation: Gifts and public goods in cyberspace, in Marc Smith and Peter Kollock eds. Communities in Cyberspace, Routledge, London.

Krishnan, R., Smith, M. D., Tang, Z., and Telang, R. 2004. The virtual commons: why free-riding can be tolerated in file sharing networks, SSRN working paper, available at SSRN: http://ssrn.com/abstract=450241.

Mirkovic, J., Dietrich, J. S., Dittrich, D., and Reiher, P. 2005. Internet Denial of Service: Attack and Defense Mechanisms. Prentice Hall PTR, Indianapolis, IN.

Monge, P. R. and Contractor, N. 2003. Theories of Communication Networks, Oxford University Press, USA.

Murphy, D. 2000. Recent "Denial of Service" attacks cost $1.2 billion, InsiderReports.com, http://www.insiderreports.com/storypage.asp_Q_ChanID_E_WB_A_StoryID_E_20000526

Naghshineh, M., Schwartz, M., and Acampora, A.S. 1996. Issues in wireless access broadband networks, in: Wireless Information Network, ed. J.M. Holtzman (Kluwer Academic, 1996).

Naraine, R. 2002. Massive DDoS Attack Hit DNS Root Servers. Internetnews.com (Oct. 23, 2002), Published on the Web, http://www.internetnews.com/dev-news/article.php/1486981.

Ng, T. S. E., and Zhang, H. 2002. Predicting Internet Network Distance with Coordinates-Based Approaches. In Proceedings of IEEE INFOCOM 2002, New York, NY, June 2002.

Nonnecke, B., Andrews, D., and Preece, J. 2006. Non-public and public online community participation: Needs, attitudes and behavior, Electronic Commerce Research (6), pp. 7-20.

Norton, W.B. 2001. A business case for ISP Peering, Published on the Web, http://www.cs.berkeley.edu/~randy/Courses/cs294.s02/Case4Peering.pdf

Odlyzko, A. 2001. Internet pricing and the history of communications, Draft, http://dtc.umn.edu/~odlyzko/doc/history.communications1b.pdf

Raymond, E. 2000. The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly and Associates, Sebastopol, California.

Saltzer, J. H., Reed, D. P., and Clark, D. D. 1984. End-to-end arguments in system design. ACM Transactions on Computer Systems 2, 277-288.

Samuelson, L. 1997. Evolutionary games and equilibrium selection, Cambridge, Mass.

Sherriff, L. 2000. Virus launches DDoS for mobile phones, http://www.theregister.co.uk/content/1/12394.html

Stearns, B. 2001. Mobile Internet and applications, Research report (June 2001).

Stahl, D.O., and Whinston, A.B. 1994. A General Economic Equilibrium Model of Distributed Computing. In New Directions in Computational Economics, Kluwer Academic Publishers, London, UK, 175-189.

Strategy Analytics, Strategy Analytics forecast $700B wireless market by 2005, http://www.strategyanalytics.com/press/PRDK007.htm

Taylor, P. and Jonker, L. 1978. Evolutionary stable strategies and game dynamics, Mathematical Biosciences (40), pp. 145-156.

Torvalds, L. 1998. What motivates free developers? Interview in First Monday (33), available at: http://firstmonday.org/issues/issue3_3/torvalds/index.html. Last retrieved on July 10, 2006.

Varian, H.R. 2000. Economic scene: Liability for Net vandalism should rest with those that can best manage the risk, the New York Times (June 1, 2000).

Wang, L., Pai, V., and Peterson, L. 2002. The Effectiveness of Request Redirection on CDN Robustness. In Proceedings of the 5th Symposium on Operating System Design and Implementation, Boston, MA, December 2002, 345-360.

Wang, X., and Reiter, M. K. 2004. Mitigating bandwidth-exhaustion attacks using congestion puzzles. In Proceedings of the 11th ACM Conference on Computer and Communications Security, Washington, DC, October 2004, 257-267.

Weibull, J. W. 1995. Evolutionary Game Theory, Cambridge, Mass.

Xiang, Y., Zhou, W., and Chowdhury, M. 2004. A Survey of Active and Passive Defence Mechanisms against DDoS Attacks. Technical Report, TR C04/02. School of Information Technology, Deakin University, Australia, March 2004.

Yavatkar R. et al., 2000. A framework for policy based admission control, IETF RFC 2753 (January 2000).

Vita

Yun Huang was born in Beijing, P. R. China on October 24, 1973, the daughter of Rongsheng Huang and Xiulian Yin. He completed her schooling in Beijing, P. R. China in 1992. In 1997 he received an undergraduate degree in Computer Science from Tsinghua University, Beijing, P. R. China. Also from Tsinghua University, he received his master's degree in Computer Science in 1999. In August 2000 he entered the Graduate School of the University of Texas at Austin in the department of Management Science and Information Systems and worked as and a research assistant at the Center for Research in Electronic Commerce. His research interests cover mobile commerce, software industry, social computing, and Internet security and competition.

Permanent address: 1-11-A-1602 Fang Cheng Yuan, Fang Zhuang, Beijing, P.R. China, 10078

This dissertation was typed by Yun Huang.